

ESTABLISHING CONTENT VALIDITY OF AN EVALUATION RUBRIC FOR
MOBILE TECHNOLOGY APPLICATIONS UTILIZNG THE DELPHI METHOD

by

Harry Walker

A dissertation submitted to Johns Hopkins University in conformity with the requirements
for the degree of Doctor of Education

Baltimore, Maryland

November 2013

© 2013 Harry Walker
All Rights Reserved

Abstract

The purpose of this research study was to establish content validity for an evaluation tool designed to measure the quality of mobile technology applications (Apps) for use in education settings. The rubric evaluation tool was developed by the researcher based on a review of the literature and consultation with recognized experts in the use of mobile technologies in education. This Delphi study was conducted in collaboration with over 90 Subject Matter Experts (SMEs) from around the world who provided feedback electronically on the domains and score descriptors that comprise the tool developed for this investigation, The Evaluation Rubric for Mobile Apps. The findings resulted in strong content validity being established for the Evaluation Rubric for Mobile Apps. Data from participants were used to refine the domains and score descriptors resulting in an empirically validated, robust evaluation tool for educators to employ in their decision making processes related to the use of mobile technology Apps in education settings. At the school and district level, this rubric has implications to ensure that limited funds available for technology purchases are used in the most effective and efficient manner. On a broader scale, researchers examining technology Apps in schools can employ the rubric in empirical studies to examine the impact of using high quality Apps on teaching and learning.

Table of Contents

Abstract	ii
Chapter One.....	1
Introduction	1
Background.....	2
Purpose of the Study.....	7
Importance of the Study.....	7
Research Questions.....	8
Summary.....	8
Operational definition of key terms.....	10
Chapter Two.....	12
Technology Use in Schools.....	12
History of Apps.....	20
Twenty-first Century Learners.....	21
Current State of Mobile Technology Integration in Schools.....	22
Current State of the Mobile App Marketplace.....	26
The Evaluation Rubric for Mobile Apps.....	29
Scoring Rubrics.....	37
The Delphi Method	39
Content Validity.....	42
Lawshe’s Content Validity Ratio	43
Summary.....	44
Chapter Three	46
Method.....	46
Participants.....	46
Setting.....	48
Instruments.....	48
The Evaluation Rubric for Mobile Apps.....	48
Validation Instruments.....	52
Design.....	54
Procedure.....	55
Statistical Analysis.....	58
Chapter Four.....	61
Results.....	61
Research Question 1	61
Research Question 2	71
Research Question 3	75
Summary.....	78
Chapter Five.....	80
Findings and Synthesis of Research Questions.....	81
Implications of the Study.....	94
Limitations.....	96
Recommendations for Future Research.....	98
References	101
Appendices	117

Appendix A, Sample of Delphi Research.....	117
Appendix B, Demographic Data of SMEs.....	118
Appendix C, Evaluation Rubric for Mobile Apps (original).....	120
Appendix D, App Rubric Online Quantitative Delphi Survey-Round 1.....	124
Appendix E, App Rubric Online Quantitative Delphi Survey-Round 2.....	125
Appendix F, App Rubric Online Quantitative Delphi Survey-Follow-up.....	128
Appendix G, Qualitative Interview with Representative Sample of SMEs.....	130
Appendix H, Rubric Adoption Prior to Delphi Study.....	131
Appendix I, Validated Evaluation Rubric for Mobile Apps.....	132
Appendix J, Website Currently Using the App Rubric.....	133

List of Tables

Table 1. <i>Subject Matter Expert Participant Pool</i>	52
Table 2. <i>Steps in the Development of the Rubric</i>	60
Table 3. <i>Timeline for Delphi Method Steps</i>	63
Table 4. <i>Minimum Values of Content Validity Ratio, One Tailed Test, $p = .05$</i>	65
Table 5. <i>Subject Matter Expert Domain Ratings from Round One Survey</i>	67
Table 6. <i>Subject Matter Expert Domain Ratings from Round Two Survey:</i> <i>Good +Excellent Measure</i>	69
Table 7. <i>Changes in Subject Matter Expert Domain Ratings from Round One to Round</i> <i>Two Survey</i>	69
Table 8. <i>Content Validity Ratio Calculations for Domain Ratings from Round One</i> <i>Survey (Good + Excellent Measure of App Quality</i>	70
Table 9. <i>Subject Matter Expert Score Point Descriptor Ratings from Round One</i> <i>Survey</i>	72
Table 10. <i>Subject Matter Expert Score Point Descriptor Ratings from Round One</i> <i>Survey: Adequately + Strongly Differentiate</i>	73
Table 11. <i>Original and Revised Score Point Descriptors for Curriculum Connection and</i> <i>Feedback Domains</i>	75
Table 12. <i>Which Score Point Descriptors Better Differentiate Levels of Quality Within</i> <i>the Domain</i>	76
Table 13. <i>Education Fields of SMEs in Follow-up Survey</i>	77
Table 14. <i>Rubric Use by SMEs to Evaluate App Quality</i>	78
Table 15. <i>Importance of Domain in Determining App Quality</i>	80

Chapter One

The purpose of this study was to establish content validity for an evaluation tool, the Evaluation Rubric for Mobile Apps. The rubric developed for this study by this researcher was designed to provide educators with a framework to make empirically based judgments about the quality of mobile technology Apps to be used in education settings. The rubric was originally posted on the researcher's educational blog in 2010 at the onset of this investigation. Chapter One addresses background related to mobile technology use and mobile technology Applications (Apps), introduces the problem, and articulates the need for a consistent and systematic approach for evaluating the potential impact of using Apps in education settings. Research questions are presented at the end of Chapter One.

Introduction

Mobile and cloud technologies are the most recent innovations in a rapidly changing and digitally connected world. The impact of these technologies will undoubtedly continue to expand in the years ahead. Rapid advances in the speed and processing power of mobile devices provide a platform for the development and use of instructional Apps in education settings. As more and more Apps emerge on the market with the potential use by teachers and students, decisions regarding their quality and parameters of use in education need to be made quickly, effectively, and judiciously.

Currently, decisions regarding which Apps to use in school settings are being made without a consistent framework to evaluate their quality and hence, potential impact on teaching and learning. Best practices in the use of mobile technologies in general, and, more specifically, effective App use in educational settings has yet to be defined and validated in the research.

In the world outside of the schoolhouse, one-to-one mobile computing in the form of “smart phones” is revolutionizing how we communicate and access information. While emerging technologies will continue to have a profound impact on our lives, the question for school leaders is, “How will mobile technology enhance the educational experience of today’s teachers and learners?” The potential impact mobile Apps may play in future efforts to improve outcomes in education need to be investigated and defined in the research to support technology integration decisions made at the school level.

Background

The positive and negative impact of mobile technologies in our culture is growing at a rapid pace. In the past five years we have moved from Palm Pilots for organizing our calendars to smart phones for organizing our lives. Since 2005, ownership of mobile technology, such as smart phones, has experienced double-digit growth among children ages 4-14. Eight-five percent of adults worldwide own smart phones and 81% have at one time used them for purposes other than making a call. Fourteen percent of smart phone owners access the Internet on a typical day up from seven percent in 2007, an increase of 100 percent in just three years (NPD Group, 2008). The NPD Group (formerly National Purchase Diary) is a North American market research company founded in 1967. From a global perspective, a recent posting on the website mobiThinking (2012), reported 5.9 billion worldwide mobile subscribers, of which 1.2 billion were mobile web users.

Given the expanding growth and impact mobile technologies are having outside the schoolhouse, the impact on teaching and learning in today’s classroom has been variable. Some school districts have implemented one-to-one initiatives such as bring your own device (BYOD) initiatives, while others still employ more traditional computer hardware

such as desktops. Some of the lag in the adoption of new technologies seen in many districts can be explained by the slow nature of change. Some may be tied to resistance to change inherent in organizations. Quinn (1996) noted in large organizations, there is a cultural and systemic resistance to change. Barriers to change can come from managers, organizational culture, embedded conflict, or pressures of conformity. Another likely impediment to more systemic adoption is the difficulties schools and educators experience trying to keep pace with rapidly, and at times, radically changing technology. In many instances, by the time new technology purchases arrive in the schools they are outdated. In addition, professional development efforts have not been focused on building teacher capacity to understand and then apply newly emerging technologies.

Empirical research is needed to determine how the use of mobile technologies might enhance the educational experience students are receiving both in and out of their classrooms. The current research base, while increasing, has done little to develop and empirically validate tools for evaluating the quality of digital content. Research is also needed to identify “best practices” for integrating mobile technologies, meshing new knowledge with what is already known about effective pedagogy.

Mobile technologies have the potential to expand the impact of teaching and learning outside the classroom walls. Shuler (2009) outlines the advantages of mobile technologies in schools including the availability of “anywhere, anytime” learning, the ability to gather, access, and process information in and outside the classroom, learning in a real-world context, the power to promote and foster collaboration and communication – essential 21st century skills, and the enabling of a personal learning experience. Shuler also notes mobile technology can overcome the challenges associated with larger technologies

such as desktop computers. The relative low-cost and accessibility of mobile devices may help to advance digital equity at a faster rate than possible with more expensive laptop and desktop computer configurations typically employed in school settings. The focus of Shuler's work was with mobile devices to access information and did not examine the impact of mobile Apps.

Increased student engagement and opportunities for collaboration have been cited in the literature and anecdotally by educators as rationales for increasing the use of technology in schools. Teachers have observed that students are more motivated, spend more time using technology, collaborate and communicate more, and benefit from having a portable and readily accessible tool (Vahey & Crawford, 2002). Van't Hooft, et al. (2004) found that students reported handhelds were easy to use, fun, and a useful tool for learning. Cole and Stanton (2003) argue that classrooms with hand-held computers differ fundamentally from more traditional desktop computing environments in that users interacting with handheld computers can also interact with each other and other computing devices at the same time, thus supporting the desire that today's students have for collaborative learning. While the Cole and Stanton study cited increased student engagement and opportunities for collaboration, it did not address how the use of mobile technology specifically contributes to improved educational outcomes.

Cheung and Hew (2009) conducted a meta-analysis of research related to the use of mobile technologies in educational settings, identifying a total of 136 articles. Of that total, over two thirds ($n = 92$) were opinion papers, and two of three of these studies used descriptive research. Only 6 of the 136 studies utilized true experimental research methods. Cheung and Hew's meta-analysis clearly identifies shortcomings in the literature related to

the lack of empirical studies to support the use of mobile technologies in educational settings.

In a case study conducted by Franklin and Peng (2008), the authors used iPod Touch devices with middle school students to help them learn about algebraic equations; in particular, the concept of slope, absolute value and elimination. Comments, observations, interviews, and surveys indicated the iPod was useful in supporting math content; however the researchers did not “test” math achievement, missing an opportunity to possibly connect the use of mobile technology with improved outcomes for students’ math skills.

Looi et al. (2010) conducted a study utilizing mobile learning to promote inquiry by transforming the third grade science curriculum for delivery via mobile technologies. The authors reported on a number of prior studies that provided designs for supporting student inquiry-based learning using mobile technologies (Roschelle, 2007; Squire & Klopfer, 2007; Chen et al., 2008; Spikol et al., 2009; Vavoula et al., 2009). Looi et al. found the science examination scores of students receiving instruction using mobilized curriculum outperformed students taught the traditional curriculum. They also observed students engaged in science learning through personal and involved ways, demonstrating greater self-directed learning. The curriculum in the Looi study did not include the use of mobile Apps.

While the literature base to support the use of technology in education has increased in recent years, what is lacking is clear empirical evidence that identifies the most effective devices, the Apps that hold the greatest promise to impact student achievement, and the best practices to leverage these devices. A logical first step in expanding the literature base related to mobile technology is to examine the potential impact of mobile Apps.

Developing a system to evaluate the quality of mobile Apps is an area that has received little attention in the research. As the use of mobile technology increases in school settings, educators would benefit from consistent structures to evaluate practices for integration, in particular guidance in selecting high quality Apps in order to ensure the most effective use of limited funding for technology purchases. The large number of Apps in the marketplace presents a challenging task for anyone charged with making decisions regarding App purchases.

There are literally thousands of choices for educators to consider when selecting Apps to support the instruction of students. The most popular category of Apps being marketed on the Internet according to the website 148Apps.biz is gaming (137,664 Apps), followed by education (91,679), entertainment (76,082), lifestyle (69,537), and business (56,406). From August 8, 2011 to May 3, 2013, the number of Apps for education more than doubled, from 40,653 to 91,679. When books, reference tools, utilities, news, productivity and navigation Apps are considered, all useful in the educational arena, the number of available Apps exceeds 275,000. The challenge for educators is to identify Apps that have the greatest potential to positively impact outcomes for students.

Given the wide variety of Apps in the marketplace, an evaluation tool for measuring App quality must be broad enough to address multiple grade/age levels and content areas, yet specific enough to address the inclusion of best practices in teaching and learning. An analytic rubric is a logical choice for these purposes as it allows for a judgment of quality, specifically where each domain supports or enhances the instructional process, to be broken out and judged separately. An analytic rubric also permits each separate domain to be scored along a continuum of quality.

Purpose of the Study

The purpose of this study was to establish content validity of an evaluation rubric designed to evaluate the quality of mobile technology Apps and explore expert opinions about the use of Apps in educational environments. The rubric was developed by the researcher based on a review of the literature and consultation with recognized experts in the use of mobile technologies in education. After posting the rubric on the researcher's educational blog, hundreds of requests for permission for its use came in to the website from practitioners all over the world. The research related to the development of the rubric domains is found in Chapter Two. A description of the evaluation rubric is found in the Instrument section of Chapter Three.

Importance of the Study

Mobile technologies have impacted many facets of our life, from communication to banking and shopping, to how we access information. The impacts of these technologies will likely increase in the coming years in areas not yet imagined. Educators should assume some responsibility for teaching students how to leverage these technologies in responsible and effective ways. While educators, policy makers, and the general public support this notion, there is a compelling need for direction, specifically empirical research, to support decision-making related to how technologies should be integrated into current instructional programs and structures. Without empirical evidence to demonstrate the benefits for increased technology use, it is difficult to make the case for additional funding, as well as more systemic use of mobile technologies in classrooms.

Providing practitioners with an empirically validated tool to guide their decision making related to App use will increase the likelihood that students have access to the

highest quality mobile Apps available to support their learning. This researcher developed the Evaluation Rubric for Mobile Apps in October of 2010. It was developed based on a review of research on best practices in teaching and learning and a review of current evaluation tools and websites for Apps. Recognized experts in the field of mobile technology use in education were consulted in the initial stages of development of the rubric. After being posted on several websites, requests for permission to use the rubric came from practitioners all over the world. As early adopters who recognized the need for such an evaluation system, these practitioners' expertise was used throughout the study to refine and validate the rubric.

The rubric will provide the research community with a needed tool to be used in studies to further examine and identify best practices for mobile technology use in schools to support teaching and learning in today's classrooms.

Results of this study will contribute to a greater understanding of how to make effective and efficient decisions regarding technology integration in educational settings. Specifically, an empirically validated tool will support the informed selection of mobile Apps to enhance instructional programs. This tool may be used in subsequent research to extend the knowledge base as to the essential components, or domains of high quality Apps.

Research Questions

The following research questions provide the focus of this research study:

1. What are the key components necessary in a rubric to evaluate the quality of mobile technology Apps used in educational settings?
2. What impact does the rubric have on the decision making process in the selection of mobile technology Apps in educational settings?

3. Which aspects of the rubric help support decision making related to Apps use in educational settings?

Summary

The need for additional research to identify best practices for integrating technology in education settings is compelling. While computers and related technologies have been present in our schools for more than 30 years, we have yet to realize the potential contributions technology may have in supporting and enhancing educational programs. This study is important because the results will provide practitioners and researchers with an empirically validated tool to support decision-making, identifying those mobile Apps that hold the greatest promise in supporting teaching and learning in education settings.

Chapter Two provides a review of the literature on mobile technology in schools, the current state of the mobile App marketplace, rationale supporting the evaluation rubric design, a review of scoring rubrics, and a review of content validity as it relates to research. Chapter Three provides the methodology for this study. Research findings are presented in Chapter Four. Chapter Five presents a synthesis of research findings, as well as limitations of the study and implications and recommendations for future research.

Operational Definition of Key Terms

Apps: Apps, is short for applications designed for mobile technology devices. While Apps have been developed for both Android and iOS mobile technology platforms, Apple and iTunes remain the predominant force in the App market. While there are many different categories of Apps (e.g. gaming, productivity), the focus of this study is with Apps designed to be used in education.

Computer-assisted instruction (CAI): A method of using computer technology in teaching and learning. Instruction is presented through a computer program to a student, or the computer is the platform for an interactive and personalized learning environment.

Gaming: Gaming involves the playing of simulations designed for the purpose of solving a problem. Their main purpose is to train or educate users.

Hand-held computer: A mobile device (also known as a handheld device, handheld computer or simply handheld) is a small, hand-held computing device, typically having a display screen with touch input and/or a miniature keyboard. An example of a hand-held would be a smart phone.

iPad: The iPad is a tablet computer controlled by a multitouch display. It has been marketed primarily as a platform for audio-visual media including books, periodicals, movies, music, games, Apps and web content. It uses a virtual onscreen keyboard in lieu of a physical keyboard and can be configured with either Wi-Fi or cellular capabilities. The iPad runs on iOS, the same operating system used on Apple's iPod Touch and iPhone.

iPod touch: The iPod Touch is a portable media player, personal digital assistant, handheld game console, and Wi-Fi mobile device designed and marketed by . The iPod Touch uses a multi-touch graphical user interface and has wireless access to the iTunes Store and Apple's App Store, enabling content to be purchased and downloaded directly on the device. The iPod Touch and the iPhone share essentially the same hardware and run the same iOS operating system. The iPod Touch lacks some of the iPhone's features and associated Apps, such as access to cellular networks, GPS navigation and the built-in compass.

Mobile technology: Mobile technology is a collective term used to describe the many types of cellular communication technology. While many types of mobile operating systems (OS) are available for smart phones and tablet computers, the two systems with the greatest market share in today's market are iOS and Android.

Rubric: According to merriam-webster.com, a rubric is a guide listing specific criteria for grading or scoring academic papers, projects, or tests. In educational terms, Wikipedia defines rubric as “an assessment tool for communicating expectations of quality.” In this study, domains will be evaluated to determine how essential each is to the evaluation of the quality of a mobile technology App. Score descriptors will also be evaluated in order to ensure clear differentiation between each level of “quality.”

Chapter Two

Chapter Two begins with a literature review of technology use in educational settings, from early uses of computer-assisted instruction to the most recent mobile technologies. Characteristics of 21st century learners and the current state of mobile technology integration in schools are reviewed. The current mobile App marketplace is presented and implications for the field of education are discussed. The review continues with an examination of scoring rubrics in research, a review of the Delphi method, and a discussion of content validity. The literature review concludes with a proposed evaluation rubric to address an existing gap identified in the research literature.

Technology Use in Schools

The first use of computer technology in education dates back to the mid-1940s (Levien, 1972). The earliest use of computers in education was confined primarily to colleges and universities given the cost and size of early computing applications. According to Levien, the first computers used for educational purposes were the MARK 1, developed by Howard Aiken and Grace Hopper in 1944 at Harvard and the ENIAC (Electrical Numerical Integrator and Calculator) in 1946 at the University of Pennsylvania. Levien noted the early uses of computers in education were primarily in mathematics, science and engineering. Computing power was used as a mathematical problem-solving tool, replacing the slide rule and thus permitting students to deal more directly with problems in type and size more common in the real world.

Fifteen years later, computers began to make their way into K-12 education. In 1959, Donald Bitier at the University of Illinois began Programmed Logic for Automated Teaching Operations (PLATO). PLATO was the first large-scale project for the use of

computers in education. The several thousand-terminal systems served undergraduate education as well as elementary school reading, a community college in Urbana, and several campuses in Chicago (Office of Technology Assessment, 1982).

In 1963, Patrick Suppes and Richard Atkinson established a program of research and development on computer-assisted instruction in mathematics and reading (Suppes, 1980). They developed self-paced computer programs that utilized drill and practice to develop math and reading skills. They incorporated instructional strategies that allowed the learner to correct his or her responses through rapid feedback that permitted the students to take a more active role in their learning.

The growth of computer use in schools increased significantly in the 1970s. In 1963, only 1% of the nation's secondary schools used computers for instructional purposes. By 1974, over two million students used computers in their classes. By 1975, 55% of the schools had access and 23% were using computers to support instruction (Molnar, 1975).

Since that time, computer access in schools has rapidly expanded to the point that most classrooms across the country have computer access today. According to a report from the U.S. Department of Education (2010), 97 percent of teachers had one or more computers located in the classroom every day, while 54 percent could bring computers into the classroom. Internet access was available for 93 percent of the computers located in the classroom every day and for 96 percent of the computers that could be brought into the classroom. The ratio of students to computers in the classroom in 2009 was 5.3 to 1.

Since computers were first introduced in classrooms, researchers have sought to examine their impact on student achievement. Although mixed results have been reported over the years, overall the use of computers has shown positive effects on student learning,

ranging from slight to moderate. The following section examines the impact of computer-assisted instruction (CAI) on student achievement when implemented through different instructional models.

It is important to note the research does not support the replacement of direct instruction delivered by a teacher with computer-based digital content. Using CAI as a complete replacement for conventional teaching may seriously weaken its effectiveness (Kulik & Bangert-Drowns, 1983).

The most effective use of CAI indicated in the research has been when it is combined with traditional teaching methods. Edwards, Norton, Taylor, Weiss, and Dusseldorp (1975) reviewed studies at various educational levels for a range of subjects and found CAI was most effective when it was used as a supplement to conventional teaching. The computer may provide a new form of presentation and allow for some learning without the presence of an instructor, however the computer alone may not be as effective as the computer with an instructor (Lowe, 2001).

Kulik et al. (1983) analyzed 48 studies of individual CAI programs that employed a combination of traditional and computer based instruction. The results from final examinations indicated in 39 of the studies, students from CAI classes scored higher than students from conventional classes. In 23 of the 39 studies favoring CAI classes, the difference in exam performance between CAI and conventional classes was statistically significant. Kulik determined that the performance of CAI students was raised by .32 standard deviations relative to non-CAI students.

In a meta-analysis that examined CAI in mathematics instruction, Burns and Bozeman (1981) found that traditional instruction supplemented by CAI was more

effective than traditional instruction alone for highly achieving students and disadvantaged students at both the elementary and secondary levels. They found no significant enhancement among average-level students.

In a recent meta-analysis focused on mathematics instruction, Cheung and Slavin (2013) examined 74 studies with a total sample size of 56,886 K-12 students. Consistent with prior reviews, their findings suggested that educational technology use generally produced a positive, though modest, effect ($ES = + 0.15$) in comparison to traditional methods. However, the effects may vary by educational technology type. Among the three types of educational technology use, supplemental CAI had the largest effect with an effect size of $+ 0.18$.

While studies and meta-analyses typically show an increase in student performance on final examinations in CAI classrooms, the impact of computers on retention of learning is less favorable. Kulik et al. (1983) reported that, of five studies with follow-up examinations measuring retention over 2 to 6 months, one study reported significantly higher exam scores in non-CAI classrooms while four studies reported higher, but not statistically significant, scores by students CAI classes.

A meta-analysis of achievement effects with microcomputer use in elementary schools was conducted by Ryan (1991). The results of 40 independent studies that looked at the effects of CAI in reading and mathematics were analyzed. Ryan found that CAI raised academic achievement, on average, by 0.309 standard deviations. Ryan concluded that a typical student's score would be raised from the 50th percentile to the 62nd percentile when exposed to CAI.

Christmann and Badgett (2000) compared the academic achievement levels of college students who had classes that used traditional instruction with those of college students who had classes in which CAI was used as a supplement to traditional instruction. The authors compiled data from 26 studies and calculated an overall average effect size of 0.127. When exposed to CAI, typical student achievement moved from the 50th percentile to the 55th percentile.

One exception to the reported positive outcomes of CAI was noted in a review of the literature conducted by Fletcher-Flinn and Gravatt (1995). The authors examined studies with CAI from 1987 to 1992 and found that there is an overall favorable effect size at all grade levels comparing CAI to traditional classroom settings. Included in the meta-analysis were several studies in which the same teacher taught both the CAI and traditional versions of a course; the researchers isolated the results of these studies and found no significant differences between CAI and instruction delivered by a teacher without the use of a computer.

In summary, most research has demonstrated CAI to have a positive impact on student achievement. Consensus in the field is that CAI has been most effective when combined with traditional direct instruction provided by a teacher. Additional research is indicated to identify best practices for sustained and differentiated implementation of CAI.

There are two questions that have not been addressed in the literature. First, what is the optimal ratio of time between direct instruction and computer-based instruction that will result in the greatest impact on outcomes for students? Second, what aspects of CAI, or in the case of this investigation, Apps will have the greatest impact on student

outcomes? The following section addresses the quality of computer programs and how it may impact the effectiveness of CAI.

A partial explanation of the modest effect sizes reported in the CAI research from the 1970s through the present may have been impacted by the quality of the programs available to researchers and practitioners. Many of the computing programs that were used in early CAI were text-based, non-graphical, and consisted primarily of drill and practice activities (Becker, 1983).

Most of the research conducted with drill and practice computer programs have shown little or no advantage to more traditional flash card instruction. For example, Fuson and Brinko (1985) found similar results when comparing the effects of computer-based drill and practice software to the (non-digital) use of flash cards. In their study, 84 second- and fourth-grade students participated in either the flash card or computer groups. Both conditions' procedures included a limited number of facts that could be individualized for specific needs, provided immediate feedback on student's accuracy with the correct answer stated, and offered immediate feedback on the speed of the students' responses. When comparing students' scores on weekly timed tests, there appeared to be no significant difference in achievement.

In recent years, researchers have begun to examine different types of computer-assisted programs to determine if there are differential effects on student achievement. As computer applications have grown in sophistication, the use of games and interactive simulations, as well as multi-media Apps have increased in schools. Possible reasons for these increases can be traced to the popularity and impact gaming has had on our culture as

a whole, as well as the increased computing power available in schools to support multimedia computer programs.

In a recent meta-analysis of the cognitive and motivational effects of gaming, Wouters, van Nimwegen, van Oostendorp, and van der Spek (2013) investigated whether “serious” games were more effective in terms of learning and more motivational than conventional instruction methods. When computer games were used to supplement other instruction methods, both learning and retention were positively impacted. In a meta-analysis conducted by Vogel, Vogel, Cannon-Bowers, Bowers, Muse and Wright (2013), the authors examined the impact computer gaming and interactive simulation had on learning. They reported strong, positive effect sizes when interactive simulations and games were compared with traditional teaching methods for both cognitive gains and attitude. The effect sizes held true across people and situations.

Introducing multimedia into CAI has shown to improve learning outcomes for children at risk of literacy underachievement. Van Daal and Sandvik (2011) conducted a meta-analysis reviewing 35 studies focusing on the effect of multimedia on early literacy development. Medium to large effect sizes were found for alphabetic knowledge (average effect size 0.654), comprehension (average effect size 0.619), phonological awareness (average effect size 0.565) and vocabulary (average effect size 0.565). Small to medium effect sizes were found for nonword reading (average effect size 0.379) and concepts of print (average effect size 0.351). It was concluded that multimedia-literacy programs could be beneficial to children at risk of literacy underachievement, especially with respect to alphabetic knowledge, comprehension, phonological awareness, and vocabulary.

The quality and design of applications used in computer assisted instructional programs has been a recent focus in the research. Ke (2009) conducted a review of 89 research studies published between 1985 and 2007 that focused on the design of computer-based games and programs for learning purposes. The findings from this meta-analysis outlined a number of considerations for the design and implementation of computer-assisted instruction.

Seventeen of the 89 studies reviewed by Ke (2009) examined the instructional design of computer games. Design features focused on pedagogy, interface format, feedback, and alignment with desired learning outcomes. A common finding from these design studies was instructional support features are a necessary part of instructional computer games. The studies generally concluded that learners without instructional support in game would learn to play the game rather than learn domain-specific knowledge embedded in the game (Mandinch, 1987; Leutner, 1993). Instructional supports may include alignment of game playing and learning task, feedback, and authenticity level.

Research has begun to demonstrate the need for careful attention to the design features of computer programs in order for CAI to have the greatest impact on student achievement. Sanford, Ulicsak, Facer, and Rudd (2007) reported that teachers' facilitation played an important role in the effective use of instructional games in the classroom. In this investigation, the authors also asserted the focus should be on how games can be carefully aligned with sound pedagogical strategies or learning conditions to be beneficial. While several researchers have articulated the need to align the design of computer games and applications with sound pedagogy, a systematic approach to the design process has yet to be identified.

It is clear that the quality of computer games and Apps would improve if best practices in teaching and learning were considered in the development stage. Identification of which practices might have the greatest impact on student achievement would provide a structure for program designers. Educators would benefit from an evaluation rubric to identify the computer games, simulations, and Apps that incorporate these best pedagogical practices. Careful selection of computer Apps utilizing a systematic approach may boost the modest effective sizes reported in past research utilizing CAI to impact student outcomes.

History of Apps

One year after the iPhone was released, the Apple App Store launched its online site. When the App Store first became available to consumers in July 2008, it had 500 Apps. Over 10 million Apps were downloaded in the first weekend (Rowinski, 2012). The Android Market launched a couple months later in October and had 50 Apps to start.

The App market is continuing to expand at an exponential rate, quickly approaching one million Apps. While an average of 775 new Apps are being developed each day, little data are available to guide the development of Apps in terms of the potential benefit they may have in education settings.

Today's students have grown up in a technology-connected world. As available Apps and App use continues to expand, schools are examining ways to leverage this technology to support teaching and learning.

21st Century Learners

Today's learners present unique challenges to the structure of the traditional classroom. Even students in economically challenged schools are using mobile

technologies for problems solving and finding information in response to their inquiries. From a historical and sociological perspective, much has been written about the latest generation of “digital native” students who have grown up surrounded by and using technologies (Strauss & Howe, 1991; Howe & Strauss, 2000).

Prensky (2001) suggested that, “digital natives are not just another generational change, rather they absorb technology to such an extent that it has changed the way these natives acquire information and indeed how they think and learn”(p. 2). Digital natives prefer receiving information quickly and are adept at processing information rapidly; they prefer multi-tasking and non-linear access to information; have a low tolerance for lectures; prefer active rather than passive learning, and rely heavily on communications technologies to access information and to carry out social and professional interactions (Frاند, 2000; Prensky, 2001; Gros, 2003; Oblinger, 2003).

Soloway et al. (2001) go a step further describing this most recent generation, stating, “The kids these days are not digital kids, the digital kids were in the 90s. The kids today are mobile, and there’s a difference. Digital is the old way of thinking, mobile is the new way. As usual, adults have not caught up.” (p. 16) Using mobile devices, children have improved ways to form their own learning networks as they communicate through conversations, texting-mail and social-networking Apps with peers and teachers (Field, 2005). Allowing mobile devices into classrooms could provide the platform for which students have a clear comfort level. The challenge for educators is determining the best practices and mobile computing Apps that hold the most promise in advancing student outcomes.

In terms of how this new generation uses technology, much has been written on the subject. Gee (2008) notes many elementary school children are gamers and emerging tech-savvy digital natives. They crave engaging experiences with new technologies and they want to learn socially and collaboratively, using digital tools that allow them to participate in learning communities and produce media and knowledge. Leveraging this digital environment by identifying games or Apps to teach, reinforce, and/or extend educational concepts and skills could have a significant, positive impact on student achievement. Engaging this new generation of learners by addressing their learning preferences is an important consideration in designing empirically researched programs that more effectively integrate technology, in particular mobile technologies.

Current State of Mobile Technology Integration in Schools

A review of the literature addressing the use of mobile technologies with school-aged children is sparse. While a larger body of research has addressed general computer use with students, mobile technologies have not received the same level of attention from researchers. Efforts to identify the best practices for the use mobile Apps have commanded even less attention from researchers. A partial explanation for this gap in the literature may be the relatively “new” nature of the technology, as well as the rapid changes and advances in the technology. For example, a little more than a decade ago, educational researchers were examining the potential benefits of using Palm Pilots, now an obsolete device in the marketplace (Hennessy, 2000).

Cheung and Hew (2009) conducted a meta-analysis of research related to mobile hand held devices in school settings and found a total of 136 articles, 92 of which were opinion papers. Of the 44 remaining studies, two thirds ($n = 29$) were descriptive research,

with only 4.5 percent ($n = 2$) utilizing true experimental research methods. Over half of 44 studies used a questionnaire or interview as their data collection method, 22.5 percent ($n = 10$) used an assessment, and 20.6 percent ($n = 9$) used content analysis. In terms of research topics, only 17.5 percent ($n = 7$) of studies focused on learning outcomes, with roughly half focused on attitudes toward the devices.

Of the 44 studies referenced in the Cheung and Hew (2009) meta-analysis, only five addressed students in elementary and middle schools. Specifically, the subjects in two studies were elementary aged students, small groups of middle school students were subjects in two studies, and one study employed both elementary and middle school aged students. Of these five studies, one was descriptive, two used survey data, and three used experimental or quasi-experimental designs. Of the three studies, the longest duration of mobile technology use was a total of six hours, with the shortest being two hours.

The limitations in the research uncovered in Cheung and Hew's study suggest one area in need of additional investigation. There is a need for research to examine the potential impact of using mobile technology with students in the primary and middle grades. Much of the current research investigating the use of mobile technologies has been with older students, as well as adults. A large percentage of App development and use is focused on younger children.

In a case study conducted by Franklin and Peng (2008), the authors used iPod Touch devices with middle school students to help them learn about algebraic equations, in particular, the concept of slope, absolute value and elimination. Comments, observations, interviews, and surveys indicated the iPod was useful in supporting math content; however the researchers did not assess the benefits of the technology in promoting math

achievement. While Franklin and Peng (2008) reported Apps helped to support math instruction, the absence of empirical data, specifically connecting the use of Apps to increases in student achievement, was a shortcoming.

Looi et al. (2010) conducted a study utilizing mobile learning to promote inquiry by transforming the third grade science curriculum for delivery via mobile technologies. The authors reported on a number of prior studies that provided designs for supporting student inquiry-based learning using mobile technologies (Roschelle, 2007; Squire & Klopfer, 2007; Chen et al., 2008; Spikol et al., 2009; Vavoula et al., 2009). The investigators in this study found the science examination scores of students receiving instruction using mobilized curriculum outperformed students taught the traditional curriculum. The observed students were engaged in science learning in personal and engaged ways and demonstrated self-directed learning.

While the Looi et al. study demonstrated increased science examinations scores for students engaged in the mobilized curriculum; the focus of the experiment was on contrasting traditional instruction with inquiry-based instruction. While mobile technology comprised a part of the mobilized curriculum, it was not possible to specifically attribute these increases in student scores to the use of mobile Apps.

As evidenced in the above-cited work in the field, a number of researchers have used qualitative measures to demonstrate positive impacts of mobile technology use with students (e.g., increased engagement and motivation). The literature is lacking in terms of studies that are able to connect mobile App use with improved outcomes for teachers and students. This connection is critical and should be an important consideration when evaluating the quality of Apps.

The use of Apps on the iPad platform and tablet computers is becoming more widespread in schools today. In a recent quarterly financial report (July 2012), Apple, Inc. reported it sold more than twice as many iPads as Macs to educational institutions while also experiencing record sales of Macs during the period. Of the 17 million iPads sold worldwide during the quarter (versus 9.2 million in the same quarter last year), about 1 million were sold to U.S. education institutions (Nagel, 2012).

Educational Apps represent the second largest genre and the fastest growing in terms of App development (Steel Media Network, 2013). In an analysis of the current App market, Shuler (2012) noted that Apps for young children were the most popular age category (58%), exceeding Apps for adults by almost 20%. Adults were the second most popular age category (40%), followed by elementary (19%), and middle school (18%). Since 2009, the percentage of Apps for children in every age category rose, accompanied by a decrease in Apps for adults. The toddler/preschool age category saw the greatest growth (23%), whereas the middle school category also saw a considerable jump (14%). In 2009, nearly half of the top 100 selling Apps targeted preschool or elementary aged children. That number had increased to 72% by 2012. Given the focus on children, Shuler states this segment of the App market “should be considered an important one for developers, researchers, investors, and policy makers” (Shuler 2009).

Shuler’s recommendation has been heeded by developers and investors, but not by researchers and policy makers. There is not a large research base to support the use of Apps in education settings. With so many Apps designated as educational, empirical research is needed to identify the most promising App qualities, as well as best practices in App use.

Given the impact mobile technologies have had in our lives, intuitively we sense they can have an impact in the educational experiences for students as well. The dilemma faced by researchers and educators is establishing empirical evidence to support their use in schools. Trying to evaluate the sheer number of Apps in the App marketplace that have been designed to support student learning compounds this challenge. Evaluating App quality, and hence their potential impact to support teaching and learning, should take place prior to their direct use with students.

Current State of the Mobile App Marketplace

In terms of the mobile App marketplace, Apple has become the predominant provider of Apps. According to 148Apps.biz (n. d.), a website devoted to iOS development news and information, there are currently 566,165 Apps, developed by 101,764 active publishers, available through the Apple Store. On average, 775 new Apps are submitted to Apple by developers every day. Elmer-DeWitt (2011) reports in a post on the CNN Money website that 15 billion Apps have been downloaded from the Apple Store by iPod, iPad, and iPhone users in the past three years. The author notes that this figure equates to more than two downloads for every man, woman and child on the planet. With over 200 million current iOS users, one can only expect the release of new Apps to continue to grow at possibly an even more rapid rate.

Defining what makes an App “good” varies depending on the audience. DotNet, a website community for .NET developers focuses on the technical aspects of the App. Their criteria include stability/reliability, consistency with the platform, fast loads, no user interface hang-ups, and absence of advertisements (Delimarsky, 2011). When

evaluating Apps for educational use, these criteria are only the bare minimum; practitioners need to take a more focused look at the educational benefits for their students.

In a July post on the website The Next Web, writer Matt Brian (2011) outlined criteria App developers should consider when creating a “popular” App. His criteria include the quality of the user interface, how effectively the App plays to the strength of mobile devices, and how effectively the App adapts to a user’s needs. Of these criteria, the quality of the user interface and the App’s adaptability are especially important considerations when evaluating Apps for their educational potential.

In terms of App developers, “popularity” is an important consideration, driven by the desire to make a profit on their work. This term used by Brian raises concerns from an educational standpoint. While motivation and engagement are important considerations in designing learning activities, they are only part of the equation. For example, *Angry Birds* which is one of the most downloaded Apps in history meets the Brian’s requirements of a quality App including popularity, but lacks merit in terms of educational value.

A number of websites are dedicated to mobile technology and App reviews, providing educators with a forum for evaluating Apps. The website, I Education Apps Review (IEAR.org), developed by Scott Meech, provides a forum for almost 1,000 educators, administrators and App developers. IEAR was one of the first App review sites that addressed the evaluation of education Apps. In Meech’s own words:

IEAR.org was my attempt to start making sense out of the App revolution and how Apps and mobile learning may or may not make a difference in the classroom. When I first began doing reviews, I was overwhelmed by the sheer

number of "bad Apps" under the educational category. I still am to be honest (Meech, 2011).

Although the IEAR community provides overviews and useful insights into the pros and cons of specific Apps, it does not provide a common language or structure for evaluation purposes. One community reviewer may rate an App as highly effective, while another reviewer may be considerably less impressed with its quality. Factors that influence a reviewer's perspective might include the targeted audience, cost, their preferred learning style, and their specific purpose for using the App. With limited funding for technology purchases, as well as the need to connect App use with increased student achievement, this lack of a common language hinders the objective reviews needed to increase the likelihood of the highest quality Apps being purchased and used with students.

Tony Vincent, an independent consultant and a national leader on the use of mobile technology in schools, has explored the issue of quality and Apps. His website, Learning in Hand, is one of the most popular resources for educators using mobile technology in schools. Vincent (personal communication, October 9, 2010) identified the following criteria as important to consider when examining the quality of an App: Usefulness, curriculum connections, ability to export/import, potential for collaboration, aesthetics, and stability. Vincent's website, Learning in Hand, does not offer a system for evaluating the quality of mobile computing Apps.

As the App market has grown exponentially, so have the number of App reviewers and App review sites available in the Internet. Using the search term, "App review websites", 131 million results were returned with websites specifically devoted to App reviews numbering in the thousands. Given the overwhelming number of Apps available in

the marketplace, and the thousands of reviewers utilizing multiple criteria, identifying a common language and structure to evaluating Apps is sorely needed.

The Evaluation Rubric for Mobile Apps

The instrument under investigation in this study, the Evaluation Rubric for Mobile Apps, was developed utilizing the recommendations from recognized experts in the field of mobile technology Apps in education as well as a review of the literature related to best practices in teaching and learning. Six of the seven domains developed for the rubric were based on recognized best practices in teaching and learning: Curriculum Connections, Authenticity, Feedback, Differentiation, Student Use, and Student Performance. The seventh, user friendliness, was included because CAI or, in recent years, App use in educational settings is usually an independent student activity without the direct support of a teacher.

The following section presents a literature review related to each of the rubric domains, providing rationale as to why they were chosen for inclusion in the rubric. A more practical explanation of each rubric domain, as well as scoring recommendations written with the App reviewer in mind, can be found in Chapter Three, *Instruments*.

Curriculum Connections

When curriculum is closely connected across environments and disciplines, learning is strengthened. Educational researchers have found that an integrated curriculum can result in greater intellectual curiosity, improved attitude towards schooling, enhanced problem-solving skills, and higher achievement in college (Kain, 1993; Austin, Hirstein, & Walen, 1997). When considering curriculum connections in a computer assisted

instructional model, it is important that online and offline resources complement one another.

Practice is a critical component in the process of learning any new skill, or transferring information from short-term to long-term memory. The advantages provided to memory by the distribution of multiple practice or study opportunities are among the most powerful effects in memory research (Benjamin & Tullis, 2010). Distributed practice of skills or knowledge can be provided by ensuring connections between online and offline curriculum. By utilizing Apps, skills and concepts introduced by the teacher during direct instruction can be practiced and reinforced.

Distributing practice of concepts and skills over time can have a positive effect on learning and retention. The literature on distributed practice is substantial. Hintzman (1974) demonstrated two spaced presentations are about twice as effective as two massed presentations and Underwood (1970) found the difference between them increases as the frequency of repetition increases. A number of researchers have conducted quantitative reviews of distributed practice (Lee & Genovese, 1988; Moss, 1996; Donovan & Radosovich, 1999; Janiszewski, Noel, & Sawyer, 2003). The authors of these reviews all noted that distributed practice results in an increase in retention. Apps may be used to provide practice of skills and concepts connected to the curriculum presented in the classroom. Distributed practice over time may have a positive impact on learning and retention.

The use of digital content in educational settings is being used to reteach, reinforce/practice, or extend learning. The issue of developing or identifying meaningful work that can be done independently can be a challenge when working with elementary-

aged children. It is important to identify high quality Apps to provide distributed practice of skills and concepts tied to the child's curriculum to increase the initial learning and retention of knowledge.

Authenticity

Learning is more relevant when it occurs in contexts where students see the value or relevance in the assigned task. Providing authentic learning experiences can ensure greater relevance for students. Unfortunately, the separation between knowing and doing has traditionally been the hallmark of school and university learning. The emphasis in school and university has been on extracting essential principles, concepts and facts, and teaching them in an abstract and decontextualized form (Resnick, 1987).

Cole (1990) contends that traditional education overemphasizes the acquisition of facts and procedures. Information stored as facts rather than as tools is 'welded' to its original occasion of use (Brown, 1997). Whitehead (1932) suggested this type of knowledge is inert, lacking power to be applied and used. These studies suggest that much of the abstract knowledge taught in schools and universities is not retrievable in real-life, problem-solving contexts, because this approach ignores the interdependence of situation and cognition.

Problem-based learning is a recent trend in education. This approach to learning recognizes that students need to apply skills and knowledge in order to truly "own" them. A number of researchers have identified characteristics of problem-based, or authentic learning including a complex, open-ended learning environment (Resnick, 1987; Collins, 1988), which reflects the way the knowledge will ultimately be applied (Collins, 1988). In terms of activities, they should have real-world relevance (Resnick, 1987; Jonassen, 1991;

Winn, 1993; Young, 1993), an authentic context and task (Norman, 1993), and where appropriate can be integrated across subject areas (Bransford et al., 1990; Jonassen, 1991).

While it presents a challenge for App developers to build this type of structure into mobile Apps, ignoring the need for authenticity in the tasks or activities of an App presents a potential risk to student engagement and motivation. Authenticity should be an important component of learning in any environment, be it digital or a traditional classroom.

Authenticity in the tasks incorporated into Apps can provide greater relevance, resulting in increased student motivation and engagement. Educational researchers have found that students involved in authentic learning are motivated to persevere despite initial disorientation or frustration (Herrington, Oliver & Reeves, 2003).

Most importantly, authenticity can result in deeper learning. The likelihood students will be able to use and apply their learning is enhanced when that learning takes place in authentic contexts. Newman, Seceda and Wehlage (1995) noted that the absence of meaning, or authenticity, inhibits learning transfer.

Feedback

Effective feedback is essential to improved performance in any learning task. Helping students to improve their performance is one of the key responsibilities of teachers. Carefully designed feedback is an important consideration while students are working on mobile Apps. Given the sense of urgency and pressure on schools today, increasing student achievement cannot be left to chance, or to trial and error. This is a solid research base that ties effective feedback to better outcomes for student learning.

John Hattie (2009) identified feedback as the single most powerful educational tool available for improving student performance. In order to make the most effective of the

limited time available during the school day, it important that learning not be left to chance, or trial and error. Little improvement in student performance can be expected in the absence of constructive feedback.

A number of researchers have identified specific aspects of feedback that make it effective in improving performance. Marzano, Pickering, and Pollock (2001) reported feedback produces the best results when it is specific, delivered frequently, and provides sufficient information on what needs improvement and how to improve. Several researchers have also connected the immediacy of results with better results (Scheeler, Ruhl, & McAfee, 2004; Coddington, Feinberg, Dunn, & Pace, 2005).

The effectiveness of feedback provided to the user while using an App is an important consideration in judging its quality. In order to be effective, the feedback should be connected to better outcomes in student performance.

Differentiation

Extensive research has been conducted related to practices of differentiation and how they can maximize learning. One way of conceiving differentiation proposed by Tomlinson (1999) is modification of teaching and learning routines to address a broad range of learners' readiness levels, interests, and modes of learning.

Researchers have identified the following effective differentiation practices: use of effective classroom management procedures; promoting student engagement and motivation; assessing student readiness; responding to learning styles; grouping students for instruction; and teaching to the student's zone of proximal development. (Vygotsky, 1978; Ellis & Worthington, 1994; Allan & Tomlinson, 2000).

The term, zone of proximal development, refers to a point of required mastery where a child cannot successfully function alone, but can succeed with scaffolding or support (Tomlinson, 2003). A challenge presented to teachers in academically diverse classrooms is the assignment of meaningful work for students to do independently following direct instruction. High quality Apps have the capability to personalize a students' learning path by assessing readiness, providing effective feedback, and determining the appropriate level of challenge for the student. Using Apps of high quality, with appropriate scaffolds and support built in, can target a student's zone of proximal development.

Kulik, et al (1991) noted student gains are greatest when instructional materials are varied for differing instructional groups, rather than using the same materials for all groups. In an academically diverse classrooms, this presents yet another challenge for the teacher. Using Apps in conjunction with the teacher's instruction can provide students with a wide variety of games, simulations, and Apps to practice and extend their learning.

A number of studies have noted the ineffectiveness of classrooms in which teachers fail to adapt the pace of instruction in response to learners' needs (Dahloff, 1971; Oakes, 1985). Often the teacher's instruction is directed to the middle in an effort to cover the curriculum. This approach may be too challenging for lower achieving students and not challenging enough for those at higher levels of achievement. Effective utilization of Apps can reduce the size of instructional groups for teachers permitting the delivery of more targeted instruction.

User Friendliness

User Friendliness is sometimes referred to in the literature as "usability", the ease of use or how intuitive something is to learn. Nielsen and Hackos (1993) note that usability

consists of a number of attributes including learnability, efficiency, memorability, errors, and satisfaction. Learnability refers to how quickly the user can learn how to use the program. Efficiency is a measure of productivity for the user once the system is learned. Memorability means the user can return to the program without have to relearn its use. Errors concern the error rate in the system, not of the user. Satisfaction is a measure of how much the user likes the program. When all of these attributes are met, the result is a positive experience for the user.

In today's digital classrooms, students may be assigned reinforcement or extension activities using mobile Apps when the teacher is not directly instructing them. In these circumstances, usability can contribute to a learning environment that functions smoothly. Intuitive devices and content increase the likelihood students will be able to work independently, freeing up the teacher to provide direct instruction to other students.

Through literature review, surveys, and response tracking over product generations, usability engineers have been able to define at least some of the components of what Risden, Hanna, and Kanerva (1997) termed a fun product. Factor analyses of children's responses to questions assessing liking and usability of computer software revealed dimensions of engagement such as "familiarity," "control," and "challenge" that fit with research and theoretical discussions of others (Malone, 1980; Lepper, 1988; Whalen & Csikszentmihalyi, 1991). This research demonstrates that ease of use is a critical determinant of engagement, and as such is key to every child's product if it is to be a success. This conclusion holds true for mobile Apps as well.

There are a number of variables that need to be considered when assessing usability. These variables include design considerations such as activity and screen design, the use of

instructions, design icons, cursor design, text-to-speech functionality, rollover features, and help screens. Each of these factors can be considered in determining the quality of an App.

Motivation

Motivation can either enhance or serve as serious impediment to student learning. There is a substantial research base connecting motivation with increased performance. When students are motivated, it increases the effort and energy they expend in activities directly related to their needs and goals, determining whether they pursue a task with enthusiasm or with apathy (Maehr, 1984; Csikszentmihalyi & Nakamura, 1989; Pintrich et al., 1993).

Researchers have noted how motivation increases students' initiation and persistence in activities, even when faced with occasional frustration (Maehr, 1984; Wigfield, 1994; Larson, 2000). When students are motivated it increases their time on task (Brophy, 1988; Wigfield, 1994; Larson, 2000). Other researchers have demonstrated a connection between motivation and improved cognitive process, specifically what learners pay attention and how effectively they process it (Eccles & Wigfield, 1985; Pintrich & Schunk, 2002; Pugh & Bergin, 2006).

Motivation is clearly an important consideration in the development and assignment of any instructional task, including the use of mobile Apps with students. A number of factors, both intrinsic and extrinsic can affect levels of student motivation. While some students will complete assigned tasks with the goal of task completion and possible resulting rewards (e.g., grades or badges), ideally the task itself, or in the case of this study, the App, should be intrinsically motivating and one a student will choose in the absence of teacher direction.

Student Performance

One of the primary purposes of education is the acquisition of knowledge. With a large number of schools failing to meet the expectations outlined in the Common Core State Standards, efforts need to be made to maximize the instructional time in the school day.

Digital content in the form of mobile Apps may help to increase the effectiveness of instructional time when they are selected and implemented effectively in a blended learning program. In any initiative or program designed to improve the quality of the education provided to children, the most important factor in terms of evaluating efficacy is the impact they have on learning.

App use has the potential to enhance teaching and learning. In order to increase the likelihood that Apps will have a positive impact in education settings, it is important to identify Apps of high quality. Defining high quality Apps has not been addressed in the literature. The Evaluation Rubric for Mobile Apps used in this study was developed based on merging a number of well-established principles of effective pedagogy with effective design principles related to technology.

Scoring Rubrics

The use of scoring rubrics has been applied in a variety of education settings for a wide range of purposes. With the advent of high stakes testing programs across the nation, rubrics have been used with even greater frequency. In school settings, rubrics have typically been used when student performance cannot be judged with objectivity, such as with constructed or essay response items on informal or standardized tests. In cases such as

constructed responses and project-based learning, rubrics are considered an effective approach for achieving valid and reliable professional judgment (Pellegrino et al., 2001).

Rubrics can be designed to use either holistic or analytical scoring. Arter and McTighe (2001) note an analytical rubric divides a product into essential traits or dimensions so that they can be judged separately. In a meta-analysis of 75 studies on the use of rubrics, Jonsson and Svingby (2007) reported rubrics can be effective tools for assessment, especially if they are analytic, topic-specific, and complemented with exemplars and/or rater training. To judge a program, or in the case of this study, a mobile computing App, an analytical rubric was the most logical choice.

A rubric has three essential features, (a) evaluation criteria, (b) quality definitions, and (c) scoring strategy (Popham, 1997). In the evaluation of a mobile computing App, evaluation criteria are the factors that need to be considered in determining the quality of the App. Quality definitions are detailed explanations that differentiate the extent by which the App meets the ideal for a particular criterion, or domain. The scoring strategy used in analytical rubrics permit for the scoring of each criterion, as well as providing a total for all the criteria. To evaluate Apps, this scoring method allows an evaluator to determine relative strengths and weaknesses across each criterion, as well as a total score for Apps targeting similar skills, concepts, or needs.

In terms of the number of levels needed in a rubric's quality definitions, there is no consensus on the ideal number. Popham (1997) suggests three to five, while Arter and McTighe (2001) suggest as many as eleven depending on the task or product being judged. There is broad agreement that the levels should be few and meaningful and the language easily understood by users (Malini, 2011).

The research base supporting the use of mobile technologies is lacking in the field of education. Particularly critical, in light of the recent expansion of mobile technology and Apps, is the need for a method to evaluate the quality of Apps and their possible use in educational settings. In a review of the literature, Allen and Knight (2009) noted research was notably sparse on the collaborative process of developing and validating a rubric that integrates data collected from academics and professionals. A validated, analytic scoring rubric would provide an iterative and sound process educators could use when making technology-purchasing decisions.

The Delphi Method

The Delphi method is an iterative process to collect and synthesize the anonymous judgments of experts using a series of data collection and analysis techniques interspersed with feedback (Skulmoski, Hartman & Krahn, 2007). SMEs are enlisted in this research method to provide feedback through surveys or questionnaires to better understand problems, to identify opportunities or solutions, or to develop forecasts. The surveys for each round are developed based on the results of the previous rounds. The process continues until the research question is answered.

The Delphi Method can be traced back to the early 1960s and has been used diverse fields such as education, government, medicine, environmental studies, and community health. Experts in these fields have been enlisted to participate in Delphi research to share their insights into future trends, to craft policies, and to identify best practices in their respective areas of expertise.

Rowe and Wright (1999) identified four key features of the Delphi method: anonymity of the Delphi participants allowing for free expression of opinions; iteration

allowing for participants to refine their views in light of other's thinking; controlled feedback informing participants of other's perspectives; and statistical aggregation of group response allowing for quantitative analysis and interpretation of data. Westbrook (1997) noted anonymity of responses allowed consensus to take place without the undue influence of rank, power, personality or persuasive speaking which is common to group meetings.

Rowe et al. (1991) suggest that the Delphi method should provide more accurate judgments than those obtained by groups that interact with each other directly. They point out the actions of a few may inhibit creativity, restrict problem solving and result in what they authors term "process loss." Murphy et al. (1998) noted that informal methods of reaching consensus, such as committees, are prone to domination by powerful individuals. The Delphi method has been described as an efficient way to gather knowledge from a group of experts (Everett, 1993; Lindeman, 1975).

There is an extensive literature base with information on methods to collect data from multiple sources. The Delphi method has been widely used to generate forecasts in technology, education, and other fields (Cornish, 1977). According to Helmer (1977), the Delphi method facilitates the formation of group judgment and has been used in education and in professional fields. The method is founded on the principle of continuous feedback and improvement.

The Delphi method has been used extensively in doctoral and master's research projects. In a review of the Delphi Method by Skulmoski et al. (2007), a search through ProQuest Digital Dissertations revealed at least 280 dissertations and theses that used the Delphi method in their research. Skulmoski includes a summary of 40 doctoral dissertations using the Delphi method. The majority of these research projects were from

the fields of education and healthcare. These studies included a wide variety of research seeking to establish consensus on future forecasting, prioritizing content, evaluating competencies, comparing concepts, developing objectives, and constructing curricula. An illustrative sample of the range of Delphi research summarized by Skulmoski et al (2007) is presented in Appendix A.

Research is notably sparse on the collaborative process of developing and validating a rubric. In this review of the literature, one single study that specifically addressed the development and validation of a rubric in the field of education was identified.

Allen and Knight (2009) conducted an investigation to formulate and test a rubric as a teaching and learning protocol for a multi-section course taught by various instructors, and to assure that students' learning outcomes are consistently assessed against the rubric regardless of teacher or section. The process used by the investigators included formulating the rubric, collecting data, and sequentially analyzing the techniques used to validate the rubric and to insure precision in grading papers in multiple sections of a course. The Delphi method was employed by the researchers to improve content and construct validity for their rubric.

Colton and Hatcher (2004) used a web-based Delphi research technique as a method to establish content validation for an online adult learning inventory. This study demonstrated the Internet could assist a group of diverse and geographically dispersed SMEs in establishing a content valid measurement.

While there is no steadfast rule for the number of Delphi rounds, typically, the method requires minimum of two rounds, but usually no more than four to achieve either consensus or stability (Brooks, 1979; Cochran, 1983; Dailey, 1988; Landford, 1972).

Landford reported most convergence of panel responses scores between round one and two. The Delphi method comes to a close when either consensus or stability of judgments is satisfied.

Content Validity

Content Validity is the extent to which a measurement reflects the specific intended construct of content (Carmines & Zeller, 1991), or in the case of the analytic rubric used in this study, the aspects that contribute to a high quality App. Assessment of content validity has typically relied on “Appeals to reason regarding the adequacy with which important content has been sampled and on the adequacy with which the content has been cast in the form of measurement items” (Nunnally, 1967, p.82). The most important consideration in determining the utility of the evaluation rubric in this investigation is content validity.

Prior to this investigation, over 200 practitioners in the field of education had adopted the Evaluation Rubric for Mobile Apps. This widespread interest in the rubric could be explained in two ways. First, when the rubric was developed in 2010, no other rubrics existed with the specific purpose of evaluating App quality. Second, input provided by SMEs obtained and incorporated into the design of the rubric addressed key domains in teaching and learning, as well as best practices identified by these SMEs. In essence, the rubric had face validity prior to the investigation. Anastasi and Urbina (1997) reported content validity could be built into rubrics by careful selection of which items to include. Foxcroft, Paterson, le Roux and Herbst (2004) noted that by using a panel of experts to review the selection of items the content validity of a test could be improved.

Lawshe's Content Validity Ratio (CVR)

One widely used method of measuring content validity was developed by C. H. Lawshe (1975). Lawshe's Content Validity Ratio (CVR) employed a formula using the ratings from SMEs who evaluated how essential particular skills or knowledge were to the performance of a task. Since Lawshe's original work, researchers to determine the content validity of a variety of tools and measures have used his CVR formula.

An online search yielded 628 references to Lawshe's work that was first presented at Content Validity II, a conference held at Bowling Green University. This method has been used for gauging agreement among raters or judges regarding how essential a particular item is to a particular construct. In Lawshe's original research, participants were asked to respond as to whether a skill or knowledge measured by an item was essential, useful, but not essential, or not necessary to the performance of the construct. If more than half the panelists indicated that an item was essential, that item had at least some content validity. Greater levels of content validity exist as larger numbers of panelists agree that a particular item is essential. Lawshe's formula yields values that range from +1 to -1; positive values indicate that at least half the SMEs rated the item as essential. A mean CVR calculated across items may be used as an indicator of overall content validity of the scale or instrument.

Lawshe's original research examined how effectively the Purdue Clerical Adaptability Test (PCAT) operationally defined the knowledge and skills needed of clerical workers. Using 14 experts who had been on the job as supervisors for at least three and half years, panel members were asked to determine how essential skills in each of six sections

on the test were in the performance of clerical work. Examples of essential skills included performing ordinary calculations and balancing entries.

Summary

Fifty years after computers were first introduced in schools, research related to their impact on student achievement has resulted in zero to minimal to moderate effect sizes. The minimal impact on student achievement in early studies can be partially explained by the limited sophistication of the computer applications themselves (e.g., drill and practice programs), as well as a lack of integration between online (i.e., instruction provided by a computer App) and offline (i.e., direct instruction by the teacher) teaching. In early computer assisted instructional models, educators and researchers may have had unrealistic expectations about the capabilities of computer programs to solely impact student achievement. As the field has evolved, two major questions have emerged. First, how should offline and online instruction be integrated to achieve maximum results? Second, how does one design and measure the quality or effectiveness of a computer App? It is this second question, the App design and its impact on effectiveness, this investigation sought to explore.

The App marketplace, while rich in terms of the number of Apps available to support education, lacks a formal system for evaluating the quality and the potential impact App use may have in the classroom. Based on best practices in pedagogy, an evaluation rubric was developed to provide support for decision-making regarding App use in educational settings. The evaluation rubric incorporated best practices in scoring rubrics as identified in the research.

The Delphi Method was chosen as the research method for this investigation in order to take advantage of the collective feedback from a broad range of SMEs working in a variety of education related fields from across the world. Lawshe's Content Validity Ratio was selected for the statistical analysis given its relative ease in calculating content validity and its straightforward interpretation of the content validity ratio.

There are a vast number of Apps in the marketplace today. App use in educational settings is increasing without a research base of identified best practices for their implementation to support teaching and learning. Tools are needed to guide educators through the process of selecting the highest quality Apps, integrating Apps into current instructional practices, and evaluating the impact of App use in terms of improving student outcomes.

The Evaluation Rubric for Mobile Applications was developed to provide a tool for educators and practitioners to assist with the decision-making process when selecting Apps to use in their settings. The rubric was developed based on research-based pedagogical practices and technology design principles. The worldwide adoption of the rubric prior to this investigation supports the need for such a tool in education settings (Appendix H). Establishing content validity for the rubric using the Delphi method will provide a research-based tool for educators to use when making decisions about which Apps to use in their settings.

Chapter Three

Chapter Three presents the research method of this study. This chapter provides details related to participants, as well as the instruments developed for this study. The chapter includes the rationale for selecting the method and the procedures utilized. The chapter concludes with a presentation of the statistical analysis used in the study.

Method

Participants

A pool of 208 Subject Matter Experts (SMEs) was invited to participate in the Delphi method to evaluate the content validity of the Evaluation Rubric for Mobile Apps. SMEs reported to be working in a number of educational fields in 35 states and the District of Columbia, as well as 7 countries (Australia, Canada, England, New Zealand, Qatar, Taiwan, and Thailand). SMEs self-identified as being employed in public and private school settings, colleges and universities, hospitals and therapeutic care settings, educational service agencies, educational cooperatives, departments of education, and private practices. Table 1 presents the various roles of SMEs invited to participate in the study.

The pool of participants consisted of practitioners who had requested permission to use the rubric in their respective fields. Of the 208 SMEs who received invitations, 94 completed the first survey (45.2% of potential participants) and 74 (35.6% of potential participants) completed the second online survey that investigated the first research question of this study. A total of 65 SMEs (31.3% of potential participants) responded to the follow-up survey. A representative sampling of SMEs participated in an interview,

which investigated the third research question of the study. This sample consisted of four SMEs who participated in at least one of the online surveys.

Table 1

Subject Matter Expert Participant Pool

Role	Number of SMEs
Instruction Technology Specialists/Directors	32
General Education Teachers	30
Assistive Technology Specialists/Directors	28
District Administrators	25
Speech and Language Pathologists	19
College/University Professors/Instructors	17
Consultants	13
Special Education Teachers	11
Library Media Specialists	9
Network/Systems Specialists	7
Educational Specialists	6
Occupational Therapists	5
School Administrators	4
Students	2

A large number of SMEs were invited in order to increase the content validity of the data obtained from the survey documents. Delbeq et al (1975) reported if disparate groups are involved in the Delphi, then a larger sample will likely be required and several hundred people should participate. Cochran (1983) found as panel size increases, reliability improves and error is reduced. As the sample size decreases, the standard error of measurement increases. Skulmoski et al (2007) noted a reduction in-group error as sample

size increases and the larger the group, the more convincingly the results can be said to be verified.

Competency related to knowledge and experience with the issues under investigation is the most important quality required of SMEs as outlined by Adler & Ziglio (1996). The competency of the SMEs in the participant pool was determined by their qualifications, education, and experience. Levels of expertise were identified through self-report data collected in the first round survey. Self-reported data regarding the background and experiences of the SME pool is found in Appendix B. Two other criteria outlined by Adler and Ziglio, capacity and willingness to participate and sufficient time to participate were evident by the panel members' agreement to participate in the study. SMEs were invited to participate via a personalized letter requesting their support.

Setting

Online surveys were employed to gather data and participants completed the surveys at their own convenience. Telephone calls were used to conduct the interviews with a representative sample of SMEs. Correspondences, including data collection and reporting survey results after each round, was done electronically.

Instruments

Five instruments were developed for this study: (a) the Evaluation Rubric for Mobile Apps, (b) round one online validation survey, (c) round two online validation survey, (d) follow-up online survey, and (e) follow-up interview questionnaire.

The Evaluation Rubric for Mobile Apps

This researcher developed the Evaluation Rubric for Mobile Apps in October of 2010 (Appendix C). The rubric focused on criteria determined to be of importance to

educators: 1) curriculum connections, 2) authenticity, 3) feedback, 4) differentiation, 5) user friendliness, 6) motivation, and 7) student performance. A practical explanation of each rubric domain as well as scoring recommendations written for the App reviewer is presented below.

Curriculum Connections

In a school setting, given the limited amount of time students spend in school, maximizing the use of instructional time is essential. Finding meaningful reinforcement and extension activities for children when they are not engaged in direct instruction with the teacher presents a challenge. The younger the student, the more challenging it can be to find meaningful activities that can be completed independently.

Evaluating the rubric, the reviewer is asked to determine how strongly the App relates to a targeted skill or concept from their curriculum. The stronger the curriculum connection, the higher score an App received in this domain.

Authenticity

The authenticity criterion addresses the quality of the instructional experiences provided to the user when using the App. Authenticity is defined as the extent students are engaged in genuine learning problems that help them connect new learning with their prior knowledge. Authenticity can range from the App of specific math skills in response to a real life problem, to “sim” type activities found in a number of popular games.

Evaluating and scoring the criterion of authenticity, the reviewer is asked to examine how effectively the App embedded authentic learning experiences opposed to rote or drill activities.

Feedback

In order to improve one's performance with any skill or concept, constructive and timely feedback is critical. Assuming some Apps will involve instruction and/or reinforcement of skills away from the direct supervision of the teacher, the quality of feedback provided to the child can make or break an App's effectiveness. A simple beep or buzz following incorrect responses will do little to improve performance, and in many cases can serve to decrease student motivation.

Effective feedback includes branching based on students' responses and attending to the partially correct answer, that is, feedback that serves to redirect students toward the correct response. This adaptive feedback may also include providing similar items to the one missed, simplifying the skill, or providing hints to students to improve their performance. Feedback includes the data available to the student and the teacher that summarizes the student's performance. This feedback allows students to chart their own progress over time, a powerful motivator for many children. This data can help teachers make instructional decisions based on students' progress and can contribute to the identification of other appropriate Apps to meet students' needs.

Evaluating and scoring this domain, higher scores are obtained if the App provided feedback that led to improved performance and had a data reporting system that informed instructional decisions or helped a child develop a personalized learning pathway.

Differentiation

The ability to set the level of difficulty or target specific skills for individual children increases the usefulness of the App as an instructional tool. Having control over the settings of the App to individualize instruction increases the likelihood of a students' success, which in turn increases their motivation. For example, a math App that allows the

teacher to control the number or complexity of factors or operations students are asked to manipulate in the App increases its potential use across a larger pool of students. In reading, differentiation may include modifications for children who have difficulty with the level of the text, or having the text highlighted and/or read aloud.

Evaluating and scoring this domain, the greater the flexibility built in to the App by the developers to allow for customization the higher the App would score.

User Friendliness

This criterion could also be thought of as ease of use. The level of support needed for a child to be able to use an App effectively determines user friendliness. Factors that can contribute to user friendliness include the option to have content or directions read aloud, color coding and step by step sequencing in math Apps, and student control allowing for the user to customize the level of difficulty of the App.

Evaluating and scoring this domain, the more independently a child was able to launch and use an App intuitively, the higher the level of user friendliness and therefore, the higher the App scored on the rubric.

Student Motivation

A highly rated App is of little value students are not motivated to use it. Distributed practice is needed in order to attain mastery of any concept or skill. If students are bored quickly with an App, motivation will certainly suffer. Novelty, the level of success a student experiences, and the quality of the interface all contribute to student motivation.

Evaluating and scoring this domain, a useful indicator in determining the level of student motivation is the frequency with which students choose to use an App on their own. Apps that have a high level of self-selection by students would receive a higher score.

Student Performance

Increasing student achievement drives all efforts in education. In these times of high-stakes testing and federal legislation, there is increased pressure on schools to ensure that all students achieve at high levels. Technology use in educational settings needs to be tied to better outcomes for students in order to ensure future funding for technology related purchases.

Evaluating and scoring this domain, the level of improved student performance needs to be evaluated. Higher scores would be obtained when there is a connection between App use and increased performance.

Validation Instruments

The round one online survey used in the Delphi portion of this investigation was designed to obtain feedback from SMEs in order to establish content validity of the Evaluation Rubric for Mobile Apps (Appendix D). This first section of this survey sought expert opinions on several facets of the rubric: (a) How important, or essential are each domain contained in the rubric in determining the quality of a mobile App? (b) Are there domains not contained in the rubric that should be considered for inclusion in this evaluation tool? (c) Does the language contained in the score descriptors permit for adequate differentiation between the score points? (d) Do revisions need to be made to the score descriptors to increase their clarity, with resulting increased reliability in scoring?

A second section of the round one Delphi survey sought to gather demographic data from each of the participants. This data included the participants experience with mobile

technology, their level of expertise compared to others in their field, their level of education, and the influence they held in their positions to impact decision-making about technology Apps and purchases.

Scoring of the first section of the Delphi survey, determining how essential each of the domains in the rubric were in evaluating the quality of an App, utilized a five point Likert scale (i.e., Poor measure, Not a good measure, OK Measure, Good Measure, and Excellent Measure). A four-point Likert scale was developed to evaluate how effectively the point descriptors differentiated the levels within each of the domains (Do not differentiate, Minimally differentiate, Adequately differentiate, and Strongly differentiate). Participants were asked to evaluate how effectively the labels captured the essence of each domain and how effectively the language used in each score descriptor differentiated between each of the score points. Feedback was requested specifically related to improving the rubric, either by added or deleting domains, or by revising the language in the domains or score point descriptors.

The round two online survey was constructed based on round one survey feedback from SMEs (Appendix E). Participants were presented with the compiled data from the round one survey asked if they would like to change their ratings for any of the rubric domains. Participants were also asked to provide their feedback on proposed language changes for the domains and score point descriptors.

The follow-up online survey was developed to gauge the adoption of the validated rubric and to further probe how Apps were being used in educational settings (Appendix F). This survey sought input from SMEs about which domains proved to be the most important in evaluating App quality. SME's were also asked to share how Apps were being used,

identify issues or problems encountered with App use, and forecast future App use in their settings.

A random sample of SMEs was interviewed using the fourth validation instrument developed for this study. The interviews were designed to probe both rubric adoption and App use utilizing an inductive approach (Appendix G). These questions were open ended and allowed SMEs to discuss App use in their settings, issues surrounding their choices of Apps, and the decision-making processes in place for purchasing Apps. These data were gathered to provide another level of App use and to inform potential directions for future research on this subject.

Design

The Delphi Method was chosen as the design for this study in order to collect information from experts to determine and test the rubric's specification and scoring procedures (McNamara, 1996). In the traditional Delphi Study, the time requirement for the completion of multiple rounds might negatively impact motivation for participation. Using a web-based Delphi survey in this study, the time factor was reduced significantly for participants, as well as the researcher. Feedback was provided in a timelier manner increasing the likelihood that participants would be motivated to continue through all rounds of the study.

This study combined a reactive component, as well as a traditional component of Delphi research (Salkind, 2007). As the name implies, the Reactive Delphi asked panel members to react to pre-generated items, rather than produce novel thinking on their own. The round one survey designed for use in this study asked participants to rate how essential each domain in a rubric is in the process of determining the quality of a mobile computing

App. Participants were also asked to make suggestions to improve the effectiveness of the rubric.

The traditional segment of a Delphi study asked participants to generate their own ideas, as to the domains to be included in the rubric. The traditional segment of the Delphi survey used in this study asked participants to identify content or issues that the evaluation rubric failed to address, as well as content that should not be included in the rubric.

The process used in the development of the rubric is presented in Table 2.

Table 2

Steps in the Development of the Rubric

Step	Description	Generalized Model	This Investigation
1	Review of research related to evaluation tools/rubrics for Apps	Internet and database searches	Review of research yielded no empirically validated instruments for App evaluation
2	Identification of key rubric domains	Review of research, consultation with experts	Consultation with mobile App developers, review App evaluation website
3	Development of mobile App Rubric	Review of rubric development literature	Rubric developed based on design principles for analytic rubrics

Procedure

The procedure of the Delphi method employed in this study is summarized below.

This study used a web-based Delphi method. Participant confidentiality was ensured through the use of an anonymous online survey that did not permit tracking information to connect particular responses with the participants. The names of the potential pool of participants, the only identifying information of individuals invited to participate in this study, were locked in a secure location with access provided only to the researchers. The procedure of the Delphi method employed in this study is summarized below and presented in Table 3.

- SMEs were invited to participate in this Delphi study to validate the evaluation rubric for mobile Apps. Consent forms were obtained electronically for each participant in the study. Participants were asked confirm their willingness to participate in all Delphi rounds in order to be considered for the study. Letters of invitation highlighted the expectations, as well as the benefits of participation. Potential participants were asked to include the following information: Title or position, education, years of experience in their field, training related to technology, and years of experience using Apps and mobile technology devices.
- Round one Delphi survey was distributed to all SMEs via the online survey tool, Survey Monkey.
- Round one data was compiled, analyzed, and shared with SMEs. Content validity ratios for each domain of the rubric, as well as the overall rubric, were calculated utilizing Lawshe's Content Validity Ratio (CVR)
- Round two Delphi survey was developed. Suggestions for additions, deletions, or improvements to the rubric put forth by individual SMEs were included in the second survey. Participants were asked if they wished to change their ratings,

consider additions or deletions to the rubric, or revise the language in the scoring descriptors to improve clarity.

- Round two Delphi survey was distributed to all SMEs via the online survey tool, Survey Monkey.
- Round two data was compiled, analyzed, and shared with SMEs. Content validity ratios for each domain of the rubric, as well as the overall rubric, were calculated utilizing Lawshe's Content Validity Ratio (CVR). The final validated rubric was distributed to all SMEs.
- After two rounds of Delphi surveys, SME responses were fairly stable and agreement regarding revisions to the language in the rubric domains and score descriptors were relatively high.
- A follow up survey was developed and distributed to all SMEs to gauge adoption of the rubric and probe App use in educational settings. This follow up survey was distributed to all SMEs via the online survey tool, Survey Monkey.
- Results of the follow up survey were shared with all SMEs.
- Telephone interviews were conducted with a representative sample of four SMEs to gauge the value and level of adoption of the finalized rubric and determine themes related to how mobile Apps were being used in a variety of education settings.
- Themes established through the qualitative interviews were distributed to all SMEs.

Table 3

Timeline for Delphi Method Steps

Timeline	Actions
Week One	Round one quantitative survey distributed
Week Four	Round one data analyzed, summary report prepared to include suggestions for additions, deletions or improvements to the rubric
Week Five	Round two quantitative survey and summary report distributed
Week Eight	Round two data analyzed, summary report prepared to include suggestions for additions, deletions or improvements to the rubric Content Validity Ratio (CVR) calculated for each domain and for the entire instrument
Week Nine	Round two summary report compiled. Rubric revised based on survey feedback from SMEs. Round two summary report and finalized and distributed to all SMEs
Week Ten	Follow up survey distributed to all SMEs
Week Twelve	Follow up survey data compiled and shared with SMEs
Week Thirteen	Qualitative interviews conducted with representative sample of SMEs
Week Fifteen	Final report and finalized rubric prepared and shared with all SMEs

Statistical Analysis

Primary analysis of the quantitative survey data employed the content validity ratio (CVR) formula developed by Lawshe (1975):

$$CVR = (n_e - N / 2) / (N / 2)$$

In this formula, n_e = number of SME panelists indicating "essential", N = total number of SME panelists. This formula yields values that range from +1 to -1. Positive values indicate at least half of the experts deemed the item essential. The following conclusions can be drawn from data utilizing the content validity ratio. When fewer than half of panel rates the item as essential, the CVR is a negative number. When half of the panelists rate the item as essential and half do not, the CVR is zero. When all panelists rate the item as essential the CVR is computed to be +1.0. When the number rate the item essential is more than half, but less than all, the CVR is somewhere between zero and .99.

In the current study, Lawshe's three point rating scale was expanded to five points in order to provide greater differentiation in participants' ratings. In place of the construct of job performance, SMEs evaluated how essential each of the domains in the evaluation rubric is in the evaluation of the quality of an App. Content Validity Ratios for each of the domains, as well as the entire rubric, were calculated from both the first and second round data.

Overall content validity is determined by examining the mean CVRs across items. Lawshe proposed the following table, outlining the minimum CVR values (Table 4).

Statistical analysis of data generated from the survey of a representative sample of SMEs employed comparative and thematic analysis. The researcher summarized data from the interviews and themes that emerged from the data were captured. Comparative analysis was completed by comparing and contrasting themes that emerged across the interviewees that formed the representative sample of SMEs.

Table 4

Minimum Values of Content Validity Ratio, One Tailed Test, $p = .05$

Number of Panelists	Minimum Value
5	.99
6	.99
7	.99
8	.75
9	.78
10	.62
11	.59
12	.56
13	.54
14	.51
15	.49
20	.42
25	.37
30	.33
35	.31
40	.29

Chapter Four

The purpose of this study was to establish content validity of an evaluation rubric designed to evaluate the quality of mobile technology Apps and explore expert opinions about the use of Apps in educational environments. Chapter Four presents the results of data analysis of surveys and interviews in the forms of narratives and tables. Research hypotheses and results are presented in this chapter. The three research questions investigated in this study are presented below.

Research Questions

1. What are the key components necessary in a rubric to evaluate the quality of mobile technology applications (Apps) used in educational settings?
2. What impact does the rubric have on the decision making process in the selection of mobile technology Apps in educational settings?
3. Which aspects of the rubric help support decision making related to App use in educational settings?

Results

Research Question 1

Research question 1 sought to identify the key components necessary in an analytic rubric to evaluate the quality of mobile technology Apps used in educational settings. The goal was to reach a high degree of consensus among SMEs in the field of education as to the essential domains that define a quality mobile App. Online surveys were administrated to SMEs to gather their expert opinions. This research employed the Delphi method to assess domains contained in the rubric. These domains were curriculum connections,

authenticity, feedback, differentiation, user friendliness, motivation, and student performance.

A relatively high level of consensus was obtained from the pool of SMEs as to the essential domains that comprise the construct “quality App.” Table 5 presents the data obtained from the first round survey illustrating the range of ratings across each of the domains from the SMEs.

Table 5

Subject Matter Expert Domain Ratings from Round One Survey

Domain	Poor Measure	Not a good Measure	OK Measure	Good Measure	Excellent Measure
	<i>N (%)</i>	<i>N (%)</i>	<i>N (%)</i>	<i>N (%)</i>	<i>N (%)</i>
Curriculum	0 (0)	0 (0)	5 (5.4)	34 (35.3)	55 (59.1)
Connections					
Authenticity	0 (0)	0 (0)	14 (14.9)	35 (37.2)	45 (47.9)
Feedback	0 (0)	1 (1.1)	6 (6.4)	36 (38.3)	51 (54.3)
Differentiation	0 (0)	1 (1.1)	7 (7.4)	23 (24.5)	63 (67)
User	0 (0)	1 (1.1)	7 (7.4)	29 (30.9)	57 (60.6)
Friendliness					
Student Use	1 (1.1)	4 (4.2)	7 (7.2)	30 (32)	53 (55.3)
Student	1 (1.1)	6 (6.5)	6 (6.5)	27 (29.3)	52 (56.5)
Performance					

Between 92 and 95 participants rated the importance of each of the domains in determining the quality of an App. Positive ratings of participants on any given domain (i.e., ratings of Good and Excellent measure) ranged from a low of 85.1% for Authenticity ($N =$

80) to a high of 94.4% for Curriculum Connections ($N = 89$). 89.8% of participants rated the domains positively ($N = 589$).

Negative ratings of participants on any given domain (i.e., Poor and Not a good measure) ranged from a low of 0% for Curriculum Connections and Authenticity to 7.6% for Student Performance. The mean negative score for the domains was 2.2%. Neutral ratings for the domains ranged from 5.4% for Curriculum Connections to 14.9% for Authenticity. The mean neutral score for the domains was 7.9% (52 of 656 total ratings).

Results from the second round survey indicated a high degree of stability in domain ratings from the 42 participants who continued in this segment of the study. The majority of participants did not change their ratings. A small percentage (i.e., 2.5% to 5%) indicated they would increase their ratings, but did not provide information as to their domain ratings in the first round survey. Since the survey was completed anonymously, it was not possible to review responses from the first round. None of the participants indicated they would decrease their domain ratings from the first round survey.

Table 6 presents data obtained from the first round survey, combining both the positive domain ratings (i.e., domains rated as good and excellent measures) from SMEs.

As indicated in Table 7 domain ratings from SMEs were fairly stable from the first and second round surveys. Between one and two participants indicated a desire to change their rating of any of the domains. In each case, these changes would have resulted in higher ratings. Participants who wished to change their ratings were to respond to the following prompt: “If you wish to change your rating, please fill out the following prompt – I wish to change my rating from ____ to ____.” The participants who indicated they would increase their rating during the second round survey were not able to recall how they

had ranked each domain in the first round. Therefore, the first round data was used to calculate the content validity ratios for each domain.

Table 6

Subject Matter Expert Domain Ratings from Round Two Survey: Good + Excellent Measure

Domain	Good + Excellent Measure <i>N (%)</i>
Curriculum Connections	88 (94.4)
Authenticity	80 (85.1)
Feedback	87 (92.6)
Differentiation	86 (91.5)
User Friendliness	86 (91.5)
Student Use	82 (87.3)
Student Performance	79 (85.8)

Table 7

Changes in Subject Matter Expert Domain Ratings from Round One to Round Two Survey

Domain	No Change <i>N (%)</i>	Increase Rating <i>N (%)</i>	Decrease Rating <i>N (%)</i>
Curriculum Connections	38 (95)	2 (5)	0 (0)
Authenticity	38 (95)	2 (5)	0 (0)
Feedback	38 (95)	2 (5)	0 (0)
Differentiation	38 (95)	2 (5)	0 (0)
User Friendliness	39 (97.5)	1 (2.5)	0 (0)
Student Use	39 (97.5)	1 (2.5)	0 (0)
Student Performance	39 (97.5)	1 (2.5)	0 (0)

Combining good and excellent domain ratings obtained from the first and second round surveys, the CVR formula developed by Lawshe (1975) was employed to calculate measures of content validity for each of the domains in the rubric. The Lawshe content validity ratio formula is: $CVR = (n_e - N / 2) / (N / 2)$.

Lawshe's original study used a three point rating system to evaluate how essential skills were to the performance of a particular job. The current study used an expanded five point rating system with two positive, one neutral, and two negative measures in order to provide participants with a three different options. The neutral choice was included in the scale so participants were not faced with a "forced choice." Table 8 presents the CVR for each of the domains in the rubric based on the number of SMEs who rated the domains as either a good or excellent measure of App quality. Good and Excellent ratings were combined as both were positive measures as determined by SMEs.

Table 8

Content Validity Ratio Calculations for Domain Ratings from Round One Survey (Good + Excellent Measure of App Quality)

Domain	Content Validity Ratio
Curriculum Connections	+ .89
Authenticity	+ .70
Feedback	+ .72
Differentiation	+ .83
User Friendliness	+ .83
Student Use	+ .74
Student Performance	+ .72

Content validity ratios (CVR) calculated from the round two survey data ranged from $+ .70$ to $+ .89$, indicating high levels of content validity for each domain, well above minimum values of CVR proposed by Lawshe and presented in Chapter Three. Given the relatively large number of SMEs who participated in the study, the CVR values obtained from the data are more than three times the minimum value proposed in Lawshe's original work. Based on these results, the Evaluation Rubric for Apps demonstrated strong content validity for measuring the quality of Apps.

As noted above, ratings from the participants remained fairly stable from round one to round two. This stability in responses resulted in the decision to cancel the planned third round of survey proposed in the original prospectus for this study.

In the first round survey, SMEs were asked if domains should be added or deleted from the rubric. No suggestions for deletions were received from participants. Participants in the first round survey suggested a total of 22 additions to the rubric. Of these suggestions, only four were referenced by more than one participant. These suggestions for additions were accessibility, data collection, research-based evidence, and cost. Accessibility referred to flexibility in the App for user with auditory and/or visual impairments. The participants who suggested data collection were interested in the App's capabilities to either store or transmit data of user performance. The suggestion for research-based evidence focused on whether or not Apps were based on sound pedagogical practices as supported by research. Several participants thought the cost of an App should be given consideration in determining quality. Given the few participants who suggested additions, the total pool of participants was not asked to provide feedback as to whether these domains should be added to the rubric. Discussion of these suggested additions are found in Chapter Five.

The second component of research question 1 involved an examination of the how effectively the score point descriptors within each domain differentiated the various levels of quality of an App. Data from the first round survey indicated high levels of agreement in terms of how effectively the score point descriptors differentiated between levels of quality within each of the domains. These high levels of agreement were likely the result of the score point having been developed based on input from experts and a review of the research related to effective rubric construction. Table 9 presents a summary of SMEs ratings of the score point descriptors for each domain.

Table 9

Subject Matter Expert Score Point Descriptor Ratings from Round One Survey

Domain	Do not Differentiate	Minimally Differentiate	Adequately Differentiate	Strongly Differentiate
	<i>N (%)</i>	<i>N (%)</i>	<i>N (%)</i>	<i>N (%)</i>
Curriculum Connections	1 (1.1)	6 (6.5)	47 (50.5)	39 (41.9)
Authenticity	1 (1.1)	3 (3.2)	48 (51.1)	41 (44.1)
Feedback	1 (1.1)	6 (6.5)	36 (38.7)	50 (53.8)
Differentiation	1 (1.1)	6 (6.5)	30 (32.3)	55 (59.1)
User Friendliness	1 (1.1)	4 (4.3)	39 (41.9)	49 (52.7))
Student Use	2 (2.2)	4 (4.3)	40 (43)	47 (50.5)
Student Performance	2 (2.2)	6 (6.5)	40 (43)	45 (48.4)

The data in Table 10 combines the number of SMEs in the round two survey who rated the effectiveness of differentiation for the score descriptors as either adequate or strong. In terms of how effectively language differentiated between levels of quality, the

percentage of SME agreement across the domains ranged from a relative low of 91.3% for Student Performance ($N = 85$) to a high of 95.7% for Authenticity ($N = 89$).

Both the first and second round surveys sought feedback from SMEs on the language used in the rubric, in terms of labels used for the domains, as well word choices for each of the score descriptors. First round feedback from a number of SMEs suggested the language in the domain, Student Use, be changed to “Motivation” to more accurately reflect the construct. Two thirds (66.7%) of SMEs in the second round survey indicated the revision better described the domain. The revised domain name, Motivation, was incorporated into the finalized rubric as a result of the data from the second round survey.

Table 10

Subject Matter Expert Score Point Descriptor Ratings from Round One Survey: Adequately + Strongly Differentiate

Domain	Adequately + Strongly Differentiate N (%)
Curriculum Connections	86 (92.5)
Authenticity	89 (95.7)
Feedback	86 (92.5)
Differentiation	86 (92.5)
User Friendliness	88 (94.6)
Student Use	87 (93.5)
Student Performance	85 (91.4)

Data provided from SMEs during the round one survey indicated difficulty differentiating between levels of quality within two domains in the rubric. Participants

provided specific feedback for improvements to the clarity of the language used in the score descriptors for the domains Curriculum Connection and Feedback. The round one feedback resulted in revisions to the language used to differentiate quality levels within these two domains. These proposed revisions to the language used in the score descriptors were presented to SMEs in the round two survey. Data from the round two survey indicated a preference from the majority of the SMEs for the revised score descriptors. Based on the data obtained in the round two survey, the revised score point descriptors were used in the final version of the rubric. The original and revised wording for the score point descriptors for the Curriculum Connection and Feedback domains are presented in Table 11.

For the Curriculum Connections domain, 92.8% ($N = 39$) of participants indicated the revised version of score point descriptors provided greater clarity and differentiation between point ratings. For the Feedback domain, 85.7% ($N = 36$) of participants indicated the revised version of score point descriptors provided greater clarity and differentiation between point ratings. Round two data from SMEs related to their preferences for word choice in the score point descriptors is presented in Table 12.

Findings from the round one and round two surveys indicated high levels of content validity for the Evaluation Rubric for Apps. Further improvements to the rubric were made based on the feedback from SMEs. These improvements included the revision of one domain label and revisions to the language in the score point descriptors of two of the domains.

Table 11

Original and Revised Score Point Descriptors for Domains

Curriculum Connection Domain Point Descriptors, Original Version

4	3	2	1
Skill(s) reinforced are strongly connected to the targeted skill or concept	Skill(s) reinforced are related to the targeted skill or concept	Skill(s) reinforced are prerequisite or foundation skills for the targeted skill or concept	Skill(s) reinforced in the app are not clearly connected to the targeted skill or concept

Curriculum Connection Domain Point Descriptors, Revised Version

4	3	2	1
Targeted skill or concept is directly taught through the app	Skill or concept(s) reinforced are related to the targeted skill or concept	Skill or concept(s) reinforced are prerequisite or foundation skills for the target skill or concept(s)	Skill or concept(s) are not connected to the targeted skill or concept(s)

Feedback Domain Point Descriptors, Original Version

4	3	2	1
Feedback is specific and results in improved student performance: Data is available electronically to student and teacher	Feedback is specific and results in improved student performance (may include tutorial aids)	Feedback is limited to correctness of student responses and may allow for student to try again	Feedback is limited to correctness of student responses

Feedback Domain Score Point Descriptors, Revised Version

4	3	2	1
Feedback is specific and results in improved student performance: Data is available electronically to student and teacher	Feedback is specific and results in improved student performance (may include tutorial aids)	Feedback is limited to correctness of student responses and may allow for student to try again	No feedback is provided to the student

Table 12

Which Score Point Descriptors Better Differentiate Levels of Quality Within the Domain

Domain	Original Version	Revised version
	<i>N (%)</i>	<i>N (%)</i>
Curriculum Connections	3 (7.1)	39 (92.8)
Feedback	6 (14.3)	36 (85.7)

Research Question 2

Research question 2 examined what impact an analytic rubric could have on the decision making process in the selection of mobile technology Apps in educational settings. To test the hypothesis that educators understood the need for systematic, consistent approach to evaluating mobile Apps and would adopt an empirically validated rubric for use in their decision-making, a follow up survey was developed and distributed to the pool of SMEs.

The purpose of the follow survey was to gauge the use of the Evaluation Rubric for Mobile Apps in education fields, identify which domains were most important to the identification of quality Apps, and explore several issues related to App use in a variety of education settings.

A total of 60 SMEs participated in the follow up survey administered six months after the round two survey. The education fields of the SMEs who participated in the follow up survey are presented in Table 13.

Table 13

Education Fields of SMEs in Follow-up Survey

Field	SME Participants
	<i>N (%)</i>
Technology Specialist	14 (23.3)
Teacher	12 (20)
Administrator	11 (18.3)
OT/Assistive Technology	10 (16.7)
Speech & Language Pathologist	10 (16.7)
University Instructor/Professor	2 (3.3)
Student	1 (1.7)

Note. Further specialization was reported by 4 SME participants (i.e., Assistive Technology Consultant, Assistive Technology Coordinator, Speech and Language Pathologist/Assistive Technology Specialist, and Innovation Coach)

The majority of participants (88.3%) in the follow up survey had direct experience using the rubric to evaluate the quality of Apps ($N = 53$). The number of times participants used the rubric in App evaluations ranged from one time to more than ten. Participant's rubric use is summarized in Table 14.

Table 14

Rubric Use by SMEs to Evaluate App Quality

Number of times rubric used to evaluate Apps	Number of participants
<i>N (%)</i>	
Once	1 (1.9)
2 to 5 times	30 (53.6)
5 to 10 times	14 (25.9)
More than 10 times	9 (16.7)

Feedback obtained from SMEs in the follow up survey indicated 98.3% ($N = 59$) of the participants would recommend the use of the rubric to colleagues who are interested in evaluating Apps. Thirty-seven participants (62.7%) indicated they would highly recommend and 21 (35.6%) would recommend the use of the rubric. One SME participant responded they would not recommend the use of the rubric (1.7%). No participants indicated they would recommend use of the rubric with reservation.

One question on follow up survey asked SMEs to rank order which domains were most important in the determination of a quality App. Average ratings calculated from participants' responses indicated *Curriculum Connections* was judged to be the most important domain related to App quality ($M = 2.19$, $SD = 0.87$). *Differentiation* ($M = 3.75$, $SD = 0.87$) was viewed as the next most important domain in the determination of App quality. The third through sixth rated domains were *Student Performance* ($M = 4.08$, $SD =$

0.87), *User Friendliness* ($M = 4.21$, $SD = 0.87$), *Motivation* ($M = 4.42$, $SD = 0.87$) and *Authenticity* ($M = 4.47$, $SD = 0.87$). The least important domain as judged by participants was *Feedback* ($M = 4.89$, $SD = 0.87$). These results are presented in Table 15.

In summary, 88.3% of participants had used the App rubric in their practices ($N = 53$). More importantly, 98.3% of participants indicated they would recommend use of the App rubric to colleagues who were interested in evaluating Apps.

Research Question 3

Research question 3 sought to determine the views of experts regarding the key requirements for using Apps, the challenges with App use, and future forecasting for App use in education settings. The same instruments employed to answer research question 2 (i.e., follow up survey and interviews) were used gather data regarding these questions and probe issues related to use of the Evaluation Rubric for Mobile Apps.

Results from participants who completed the follow up survey indicated App use was widespread across age groups from preschool to college. The highest level of App use by SMEs was in the primary elementary grades (76% of participants). As referenced earlier in Chapter 2, this is the fastest growing segment of the App marketplace.

Participants were asked “How are you currently using Apps in your setting?” The two content areas where Apps were being used most frequently by SMEs who responded to the survey were Language Arts (92.7%) and Math (78%). Science and Social Studies were two content areas where at least half of the participants indicated they had used Apps (51.2%). App use was also reported in Music (31.7%), Art (29.3%), and Foreign Language (26.8%). More specialized App use was noted by several SME participants. These uses

Table 15

Results of Rank Ordering of Domains in Determining App Quality

Domain	1	2	3	4	5	6	7	Rating Average
	<i>N</i> (%)	<i>N</i> (%)	<i>N</i> (%)	<i>N</i> (%)	<i>N</i> (%)	<i>N</i> (%)	<i>N</i> (%)	<i>M</i> (<i>SD</i>)
Curriculum	28 (52.8)	10 (18.9)	3 (5.7)	6 (11.3)	3 (5.7)	1 (1.9)	2 (3.8)	2.10 (0.87)
Connections								
Authenticity	5 (9.4)	6 (11.3)	4 (7.5)	8 (15.1)	13 (24.5)	8 (15.1)	9 (17.0)	4.47 (0.87)
Feedback	1 (1.9)	4 (7.5)	8 (15.1)	7 (13.2)	11 (20.8)	11 (20.8)	11 (20.8)	4.89 (0.87)
Differentiation	4 (7.5)	11 (20.8)	10 (18.9)	12 (22.6)	6 (11.3)	5 (9.4)	5 (9.4)	3.75 (0.87)
User Friendliness	5 (9.4)	8 (15.1)	7 (13.2)	7 (13.2)	8 (15.1)	13 (24.5)	5 (9.4)	4.21 (0.87)
Motivation	2 (3.8)	6 (11.3)	13 (24.5)	6 (11.3)	7 (13.2)	11 (20.8)	8 (15.1)	4.42 (0.87)
Student Performance	8 (15.1)	8 (15.1)	8 (15.1)	7 (13.2)	5 (9.4)	4 (7.5)	13 (24.5)	4.08 (0.87)

Note: Rank ordering of domains in determination of App quality – 1 most important, 7 least important

included augmentative and alternative communication, behavior, organization, self-regulation, and social skills.

Participants were asked “How are you currently using Apps in your setting?” A number of themes emerged from this open-ended question related to App use. The most widely referenced use of Apps by SMEs was to support instruction. This type of App use included reinforcement, practice, extension, and supplement to instruction focused on skills and concepts. The second most referenced use of Apps focused on communication, specifically Apps used as augmentative or alternative communication tools and as a tool to develop language skills.

A number of themes emerged when SMEs were asked, “What problems have you encountered in using Apps in your setting?” The quality of Apps, or lack thereof was the most common theme identified by SMEs. Many participants indicated difficulty finding quality apps to support teaching and learning. Quality issues included App performance, accessibility, technical and management problems, ads on free or lite App versions, problems with updating, and the inability to export data and share Apps. The second theme that emerged in survey responses was not directly related to Apps, but rather infrastructure issues in their settings. Wi-Fi and bandwidth issues were cited by a number of participants as negatively impacting App use.

The final question of the follow-up survey asked SMEs to future forecast in response to the question, “What do you see in the future for App use in your setting?” The overarching theme that emerged from the group was the consensus App use would continue to expand and improve in future years. A number of participants provided their feedback about how Apps would change/improve in the near future, including more content creation

Apps, greater collaboration when using Apps, greater customization and differentiation features, and multi-function Apps. The ability and willingness of professionals to share the best Apps was another theme that emerged. Several participants mentioned the use of the App rubric under investigation as a method to identify those Apps.

Interviews were conducted by phone with a representative sample of SMEs who participated in the survey portions of this investigation in order to gather additional data to answer research question 3. While providing additional details about App use in educational settings, similarities in themes were noted in the interviews with the responses of SMEs in the follow up survey.

Interviewees indicated the Evaluation Rubric for Mobile Apps had been a useful tool in the decision making process about which apps to purchase. Adoption of the rubric has ranged from the individual teacher level to system-wide use to determine which apps should be considered for purchase and use. All interviewees recommended the use of the rubric to colleagues.

Interviewee's feedback regarding which domains were most important to determine App quality mirrored the data obtained in the follow-up survey. Curriculum Connections were deemed to be one of the most important considerations. As one interviewee stated, "If it doesn't match the curriculum you have to ask yourself why are we using it?" One interviewee indicated use of the domains, but not necessarily the ratings, stating, "I look more at is it strong or not in each domain. No app is perfect."

The types of Apps being used were similar to the data obtained in the follow-up survey. While a number of content areas were referenced, all interviewees indicated they had evaluated Math and Language Arts Apps. Similar themes regarding App use were also

noted in the interviews. Apps were used for individuals, small groups, and as motivation for students.

Instructional implications shared by interviewees addressed the need for additional support. The lack of a systematic plan for App use was referenced by each SME. One interviewee stated, “I think we are just starting to understand how to use apps to support what we are doing in the classroom. We need help.”

Two interviewees noted a lack of funding for Apps that was consistent with many comments obtained in the follow-up survey. This would appear to be associated with the preceding point about a lack of a systematic plan to purchase and use Apps.

All interviewees supported the data obtained in the follow up survey regarding future forecasting. All stated that App use would continue to grow. One interview stated, “Mobile devices and Apps may become school supplies.” While growth was a consistent theme, two interviewees provided the following insights. One said, “I am hoping the quality of Apps improves and that designers start directly addressing things like the Common Core.” Another said, “Teachers need professional development if they are going to use Apps more effectively.”

Summary

The results of the study contributed to the research base in effective uses of mobile Apps to support teaching and learning. Using the Delphi method, a large a pool of SMEs in a variety of education fields provided feedback to refine and validate a tool designed to evaluate the quality of Apps for use in education settings. Feedback from SME participants in two rounds of online surveys was used to calculate content validity for the Evaluation Rubric for Apps. High levels of content validity were determined utilizing Lawshe’s

Content Validity Ratio for each of the domains in the rubric, as well as for the overall instrument. Data from a follow up survey of SMEs and qualitative interviews from a representative sample of SMEs provided information as to how Apps are being used in education settings, issues with App implementation, and future trends for App use to support teaching and learning. It is evident that additional research needs to be conducted in order to identify the most effective ways to incorporate App use into current practices in education settings.

Chapter Five

Chapter Five presents a summary and synthesis of the findings for the research questions. Implications and limitations of the study are presented next based on the findings of the study. Recommendations for future study conclude this chapter.

The use of mobile technologies has continued to increase in education settings. The research base to support best practices in the use of these emerging technologies has lagged behind. As more schools employ tablet computers and implement one-to-one and bring your own device (BYOD) initiatives to support teaching and learning, App use has also increased. Educational Apps represent the second largest and fastest growing genre of mobile applications. Many App review sites have emerged as the market has expanded, but a research-based system for determining App quality has not been developed. This lack of common ground, particularly in regard to what constitutes a quality App, limits research efforts to identify best practices of App use to support teaching and learning.

This research study investigated whether SMEs from various education fields could contribute to the development and validation of an instrument to evaluate App quality. The Delphi method was used as the research design. The method enabled the collection and analysis of input and feedback from ninety-four educators and practitioners who demonstrated subject matter expertise in the use of mobile technologies in education settings. The data was gathered through online surveys and telephone interviews. Both quantitative and qualitative data were collected to answer the three research questions presented in Chapter One. A review of the research questions and a summary of the results follow.

Findings of the Research Questions

1. What are the key components necessary in a rubric to evaluate the quality of mobile technology Apps used in educational settings?

Research question 1 examined whether high degrees of agreement could be established among SMEs as to the essential domains to determine the quality of Apps. The findings for research question 1 resulted in high content validity being established for the Evaluation Rubric for Mobile Apps. High content validity was also demonstrated for each domain that comprised the rubric. Feedback from participants was used to clarify the language of the score point descriptors. As a result of these revisions, 90% of SMEs indicated adequate to strong differentiation between the levels of App quality in each of the domains.

As part of the online survey, SMEs were asked whether domains needed to be added or deleted from the rubric. Accessibility was one of the suggestions for an addition to the rubric made by several participants. The concept of accessibility was already embedded in the rubric. Access for students with disabilities was included in the domain, Differentiation. The language used in the score descriptor for a “4”, or top rating of quality read, “App offers complete flexibility to alter settings to meet student needs.” In addition to being able to adjust the content and complexity of the skills, the differentiation domain included modifications for students at various levels of performance (e.g., App can read aloud to the user). To ensure users were aware of the scope of what was to be evaluated in these domains, explicit notes were added to the scoring guide for the rubric.

The second suggestion provided by at least two SMEs for an addition to the rubric was data collection. This suggestion was viewed as an omission that should have been

considered in the development of the rubric, and therefore worthy of inclusion. This decision was based on the research base to support the importance of data in the process of making instructional decisions to promote student growth. A number of researchers have suggested to gain deeper insight into students' learning needs, teachers should examine evidence from multiple data sources (Halverson, Pritchett & Watson, 2007; Herman & Gribbons, 2001; Lachat & Smith, 2005). App data, triangulated with multiple data sources, could contribute to the development of a learning profile for each student, enhancing the effectiveness of educational programming. These additional data sources could include standardized tests, as well as formal and informal assessments conducted by teachers and/or practitioners.

Rather than add an eighth domain to the rubric, data collection was addressed by revising the wording in the domain, Feedback. The highest level score descriptor, level 4, was revised to read, "Feedback is specific resulting in improved performance; Data is available electronically to student and/or teacher." The revision to this score point descriptor in the Feedback domain was included in the second round survey for feedback from SMEs.

Research-based evidence to support the decision making process of App selection was a third addition suggested by two of the SMEs. Empirical research related to App quality is not available. The impact App use may have on student achievement is another shortcoming in the literature. This suggestion was not considered for an additional domain in the rubric due to the lack of a research base to support the potential impact of App use. The results of this investigation will support future research-based efforts to identify best practices. Researchers will have an empirically validated tool to select quality Apps and to

begin the process of identifying best practices related to App use. As researchers identify best practices, data obtained from future studies may indicate the need for additions to the rubric in future iterations. Research based interventions will ensure educators and practitioners are aware of the most effective uses of Apps to support teaching and learning. Without the support of the research, App use is left to the discretion of individual practitioners who may have differing views about what constitutes a best practice.

Cost was the fourth suggestion for an addition to the rubric suggested by more than one of the participants. The issue of cost is complicated given the current state of the App marketplace. More expensive Apps may not necessarily be of higher quality than less expensive Apps, making it a challenge to identify a direct relationship between cost and quality.

The cost of Apps is driven by a variety of market forces. Competition has been one factor that has driven the cost lower for Apps. In a market of over a million Apps, this might explain why the majority of Apps are in the \$0.99 to \$2.99 price range. Another reason for the decrease in App retail prices is the availability of alternative revenue streams for developers. Rather than charge a flat, up-front fee for an App, App creators are beginning to spread revenue generation across the entire App experience, particularly in-App purchases and in-App advertising.

The level of specialization of an App can also affect its cost. Apps that are developed for special needs populations such as assistive communication Apps are generally more expensive than those designed for more general use in education settings. For example, the cost of an App developed for this purpose by AssistiveWare entitled, Prologue2Go, is \$189 in the App Store. It is difficult to correlate App quality and cost. App

cost is a complex issue driven by a number of market forces. As a result of the factors presented above, cost was not added as a domain to the rubric.

Data collected from SMEs during the round one survey indicated the score descriptors developed for two of the domains (i.e., Curriculum Connections and Feedback) did not permit adequate differentiation between the levels of quality. As a result, feedback from SMEs was used to revise the language of score descriptors in these domains. Revised score point descriptors were included in the round two survey for feedback from participants. The majority of participants (92.3%; $N=69$) indicated the revised score descriptors for the Curriculum Connections domain provide greater clarity permitting for better differentiation between levels of quality within the domain. The majority of participants (84.6%; $N=64$) also indicated the revised score point descriptors for the Feedback domain improved the clarity and differentiation between quality levels.

SME ratings of the essential components of a quality App confirmed the research that contributed to the initial development of the Evaluation Rubric for Mobile Apps presented in Chapter Two. Overall, high levels of agreement as to what comprises a quality App demonstrated that each of the rubric domains were either good or excellent measures. The following sections present a discussion of each rubric domain as it relates to research on teaching and learning.

Curriculum Connections was the highest rated domain by SMEs in both the round one survey and follow up surveys. In the round one survey, 94.4% ($N=88$) of participants indicated the Curriculum Connections domain was a good or excellent measure of App quality. Participants in the follow up survey indicated Curriculum Connections was the most important domain in determining App quality.

The findings of this study support the research that has examined the importance of curriculum connections in supporting teaching and learning. Research has demonstrated how distributed practice can promote learning and retention (Lee & Genovese, 1988; Moss, 1996; Donovan & Radosevich, 1999; Janiszewski, Noel, & Sawyer, 2003). Providing meaningful and engaging practice activities connected to the curriculum can present a challenge to teachers. Data from SMEs supported the use of Apps as an effective means to provide practice of skills and concepts connected to the curricula.

The second highest domain as rated by SMEs in the round one survey was Feedback. More than 92% ($N = 97$) of participants indicated this domain was a good or excellent measure of App quality. It is interesting to note in the follow up survey Feedback was the lowest ranked domain (i.e., average rating 4.89). A possible explanation might be a disconnect between the ideal as envisioned by SMEs in the round one survey and the reality of the current App marketplace. Many Apps that are currently available to practitioners are not able to provide users the sophisticated levels of feedback that are identified in the higher levels of the score descriptors in the Feedback domain.

The findings of this study support the research that has examined the value of feedback to support teaching and learning. John Hattie (2009) identified feedback as the single most powerful educational tool available for improving student performance. The importance of effective feedback has been reinforced by a number of other researchers (Marzano, Pickering, & Pollock, 2001; Scheeler, Ruhl, & McAfee, 2004; Coddington, Feinburg, Dunn & Pace, 2005). Data from SMEs validated the importance of providing the App user with meaningful feedback to improve their performance.

SMEs ranked Differentiation and User Friendliness at the same level of importance in the round one survey. In the round one survey, 91.5% ($N = 86$) of participants indicated both domains were good or excellent measures of App quality. These domains were rated by SMEs as being the third most important in determining App quality. Both domains were rated at nearly the same level of importance in the follow up survey. Differentiation was rated as the second most important domain and User Friendliness the fourth most important.

The findings of this study support the research that has examined the importance of differentiation to support teaching and learning. Tomlinson (1999) proposed effective differentiation should include the modification of teaching and learning routines to address a broad range of learner's readiness levels, interests, and modes of learning. A number of researchers have demonstrated the importance of targeting a student's zone of proximal development for instruction (Vgotsky, 1978; Ellis & Worthington, 1994; Allan & Tomlinson, 2000). Flexibility of an App to meet individual learners' needs was a key component in the rubric. Data from SMEs indicated differentiation was an important domain in determining the quality of an App.

The research on user friendliness, or usability, is a relatively recent development. The increased attention in the research is related to the increase in the use of computers over the last three decades. In order for the user to have a positive experience, Nielsen and Hackos (1993) found that a number of attributes were necessary including learnability, efficiency, memorability and satisfaction. User friendliness as defined in the rubric addressed the attributes of learnability and memorability. Satisfaction was incorporated into a separate domain, Motivation. Data from SMEs in all survey rounds validated the importance of user friendliness as a component of a quality App.

Student Use, later revised to Motivation based on SME feedback, was the next highest rated domain. In the round one survey, 87.3% ($N = 82$) of SMEs rated this domain as a good or excellent measure of App quality. The data from the follow up survey placed this domain in the same position (i.e., fifth most important domain) as rated in the round one survey.

The findings of this study support the research that has examined the importance of motivation to support teaching and learning. Researchers have connected motivation to increased performance (Maehr, 1984; Csikszentmihalyi & Nakamura, 1989; Pintrich et al., 1993), initiation and persistence in activities (Maehr, 1984; Wigfield, 1994; Larson, 2000), time on task (Brophy, 1988; Wigfield, 1994; Larson, 2000) and improved cognitive processes (Eccles & Wigfield, 1985; Pintrich & Schunk, 2002; Pugh & Bergin, 2006). Data from SMEs in all survey rounds validated the importance of Student Motivation as a component of a quality App.

In the round one survey, Authenticity was rated by 85.1% ($N = 80$) of SMEs as a good or excellent measure of App quality. This domain was ranked the least most important based on SME ratings in the round one survey. Authenticity was rated as the sixth most important domain by participants who filled out the follow up survey (e.g., average rating of 4.47). Given the research base that supports the value of authenticity in promoting learning, higher SME rankings for this domain would be expected. While the Content Validity Ratio for Authenticity was strong (i.e., $+ .70$), the relatively low ranking (i.e., sixth out of seven domains) was unexpected.

The relatively low ratings and ranking for the Authenticity domain warrant further discussion. One reason for this low ranking may be the limited availability of Apps that

employ a problem based approach. While researchers have demonstrated the value of PBL and educators have begun to embrace this approach to learning, the current state of the marketplace is lacking in Apps that support this approach. This finding may spark the development of more Apps that use problem-based learning as an approach to teaching and learning.

The findings of this study support the research that has examined the importance of authenticity to support teaching and learning. Much of the research base related to authentic learning has involved problem-based learning, or PBL. Researchers have demonstrated authentic learning tasks promote a student's perseverance when faced with a challenge or frustration (Herrington et al., 2003) and deeper learning (Newman, et al., 1995). Data from SMEs in all survey rounds validated the importance of Authenticity as a component of a quality App.

The lowest rated domain in the round one survey was Student Performance. Eighty-four percent of SMEs rated this as a good or excellent domain in determining App quality. While the lowest ranked of the seven domains, it was still favorably ranked (e.g., good or excellent measure) by 79 of the 94 SMEs. In the follow up survey, Student Performance was rated as the third most important of the seven domains in the determination of a quality App.

The discrepancy in ratings from the round one survey and rankings in the follow up survey might be explained by changes in opinions based on the participants' experiences using Apps. Participants may have come to realize that improved student performance should be ranked higher than the round one data indicated. In other words, if Apps are to be

used in education settings, there should be improvements in student performance to justify the use of instructional time.

In summary, data collected from SMEs during the course of the investigation resulted in high levels of content validity being established for the Evaluation Rubric for Mobile Apps. Strong content validity for the rubric was the result of research based best practices in pedagogy being incorporated in its design. Based on the relative stability of the data, only two survey rounds were required to establish the rubric's content validity.

The validated rubric contributes to the field by filling a void that to date had not been addressed in the research. Prior to this investigation, practitioners and researchers did not have an empirically validated tool to evaluate the quality of Apps. While App use in education settings has proliferated in recent years, research aimed at measuring efficacy of App use has been lacking.

The validated rubric provides a system to evaluate Apps in order to identify those of the highest quality. Practitioners across the globe have used the rubric to evaluate thousands of Apps. Appendix J is one example of a website developed by a practitioner in Australia who has reviewed over 750 Apps using the rubric. With high quality Apps identified, researchers are now able to explore a number of questions regarding the most effective use of Apps in education settings. These questions include: Which instructional models (remediation, reinforcement/practice, extension) might make the most effective use of Apps to support teaching and learning? What is the optimal amount of time for students to be engaged using Apps to promote student learning? Which types of Apps have the greatest impact on student learning?

2. What impact does the rubric have on the decision making process in the selection of mobile technology App in education settings?

Research question 2 examined the impact of the Evaluation Rubric for Mobile Apps in education fields. The follow up survey was designed and distributed to SMEs to gauge the extent the rubric has been used in actual practice. Phone interviews were conducted with a representative sample of SMEs to gather additional data to evaluate the impact of the rubric. In general, the themes that emerged from the interviews were consistent with the themes found in the data from the follow up survey.

Through their responses to interview questions, a representative sample of SMEs confirmed the value of the Evaluation Rubric for Mobile Apps in the decision making process related to App use and purchases in education settings. All interviewees had used the rubric multiple times to evaluate App quality and had continued to share the rubric with colleagues.

While all interviewees indicated personal use of the Evaluation Rubric for Mobile Apps, they noted the need for more direction and support at the school-wide or system-wide level. These early adopters are using the rubric and other evaluation websites and tools, but processes to evaluate Apps on a larger scale appear to be lacking. Administrative issues related to App use included the lack of school-based or system-wide approval and purchasing processes, questions related to who can manage accounts and download Apps, a shortage of devices for students, and issues with bandwidth. Another problem included a lack of funding for Apps resulting in educators using their own funds to buy Apps for use in their settings.

A number of themes emerged when interviewees were asked about instructional implications of App use in their settings. All of the interviewees indicated App use in their settings was being used in conjunction with instruction provided by the teacher.

Instructional approaches included small group and individual practice, reinforcement of learning, stations, and as motivation. Research has demonstrated combining CAI with traditional instruction is the most effective in promoting learning (Edwards et al., 1975; Burns & Bozeman, 1981; Kulik et al., 1983, Lowe 2001; Cheung & Slavin, 2013).

All interviewees indicated the desire for consistent plans and direction from within their agency/system. One interview indicated that much of what they are doing is hit or miss. Another said, “We need a lot of help.” Interviewees noted problems navigating the App marketplace. These problems included difficulty combing through such a large number of Apps and finding Apps that would do what they wanted them to do. While the Evaluation Rubric for Mobile Apps has helped support practitioners, interviewees indicated a need for wider support from leaders within their organizations. This theme is supported in the research. Flanagan and Jacobsen (2003) point out many principals have not been prepared for their new role as technology leaders leading to struggles to achieve the best technology outcomes in their schools. A study conducted by Anderson and Dexter (2005) found that while technology infrastructure was important, technology leadership was even more critical for the effective utilization of technology in schools.

When asked about future trends in App use, all interviewees indicated it would continue to expand in the future. One interviewee expressed hope that developers would improve the quality of Apps and being to address the Common Core. Others pointed out the need for professional development if teachers are going to use Apps more effectively.

This need for professional development was raised in the online survey and has also been addressed in the research. Flanagan and Jacobsen (2003) noted limited access to appropriate ongoing professional development was a significant impediment to teachers being able to successfully integrate technology. In a review of the literature conducted by Brinkerhoff (2006) insufficient professional development focused specifically on technology integration limited more widespread adoption of computer use in schools (Butler & Sellborn, 2003; Cuban et al. 2001; Loveless, 2003; Pelgrum, 2001). The need for professional development is indicated whenever new strategies or methodologies are introduced in education settings. The research underscores how critical professional development is to successful technology integration.

In summary, the Evaluation Rubric for Mobile Apps has already had an impact on the field as evidenced by its international adoption by individual practitioners, as well as agencies and districts. The web presence of the rubric is significant resulting in ongoing requests for permission to use rubric from around the world. Through surveys and interviews, educators and practitioners indicate the need for more systemic support from their leadership related to App use. This support included additional funding, as well as professional development in order to build capacity to best leverage mobile technology. Future forecasting from participants indicated App use will continue to expand.

3. Which aspects of the rubric help support decision making related to App use in education settings?

Research question 3 examined quantitative and qualitative data to determine if certain domains were more important than others in making decisions about App use in education settings. Qualitative data collected through open-ended surveys and interviews

identified issues surrounding app use and explored future forecasting of app use in education settings.

Quantitative data from the round one survey and follow up survey revealed consistent ratings/rankings of the domains in the Evaluation Rubric for Mobile Apps in the determination of App quality. Strong content validity was established for each of the domains that comprise the rubric. While some variance was noted, overall rankings of the most important domains obtained from the round one survey were consistent with the data from the follow up survey.

Interviewees' rankings of the most important domains of the rubric were similar to the data obtained in the online survey rounds. Of the four interviewees, two indicated Curriculum Connections were the most important domain to consider when evaluating the quality of Apps. The Curriculum Connections domain also obtained the highest Content Validity Ratio in this investigation.

Practitioners clearly recognized the need for any App use in education settings to tie to the curriculum or individual goals of students and clients. In order to develop a deeper understanding of content or develop greater proficiency with a skill, practice is a critical component. Practice is also important to the transfer of information from short-term to long-term memory. The advantages provided to memory by the distribution of multiple practice or study opportunities are among the most powerful effects in memory research (Benjamin & Tullis, 2010). By tying App use with skills presented in education settings learning can be enhanced.

A number of researchers have conducted quantitative reviews of distributed practice (Lee & Genovese, 1988; Moss, 1996; Donovan & Radosevich, 1999; Janiszewski, Noel, &

Sawyer, 2003). The authors of these reviews all noted that distributed practice results in an increase in retention. Apps may be used to provide practice of skills and concepts connected to the curriculum presented in the classroom. Distributed practice over time may have a positive impact on learning and retention.

In summary, all domains proved to have utility in determining the quality of mobile Apps. Data from the surveys indicated Curriculum Connections was the most important domain in terms of evaluating the quality of Apps.

Implications of the Study

Integrating mobile technologies into instructional models hold promise for improving outcomes for students. Research has demonstrated the most effective computer assisted instruction models have combined instruction via a computer programs with traditional instruction provided by a teacher (Edwards et al., 1975; Burns & Bozeman, 1981; Kulik et al., 1983, Lowe 2001; Cheung & Slavin, 2013). App use in education settings should follow this same instructional model until new research is conducted to confirm and/or identify best practices for App use to support teaching and learning.

Early research with CAI yielded disparate effect sizes. Lower effect sizes obtained in the 1970s may have been impacted by the quality of the computer programs. Becker (1983) noted early CAI programs were text-based, non-graphical and consisted of primarily drill and practice. Since that time, computer programs and most recently, Apps have evolved in their sophistication. There appears to be a growing awareness among developers as to the importance of developing programs, or Apps, that incorporate best practices in teaching and learning. The Evaluation Rubric for Mobile Apps can provide developers with a framework to assist in these efforts.

The lack of App quality has been clearly articulated as a concern by practitioners in education settings. A common language and system to evaluate Apps is clearly needed to guide practitioners in their decision making related to App selection and use. Educators are faced with thousands of Apps choose from and little support to guide their decision-making. Given limited instructional time and funding for technology, this research-based evaluation tool can help to support this decision-making.

In order to have the greatest impact on the effectiveness of mobile technology Apps in education settings, practitioners should take a more active role in the App development process. As “consumers” of this technology, they are in a position to effect change. A key to improving the quality of Apps might involve a partnering between practitioners and App developers. Clearly articulated expectations from practitioners in terms of the criteria needed of Apps to be used in education settings will help to drive the development of new Apps in the market. Using tools such as the validated Evaluation Rubric for Mobile Apps to connect effective design principles with research related to teaching and learning will lead to improvements in the quality of education Apps.

Another implication of this study involves the role of leadership within agencies and school systems. As early adopters to mobile technologies, it is apparent the SMEs who participated in this study are seeking support from their leadership teams. In many cases, these individuals and small groups are networking and problem solving on their own. Leadership teams in school systems and agencies need to take more active role as one-to-one and BYOD initiatives continue to expand. Research-based best practices need to be adopted and systemic policies and processes developed to ensure the most effective use of Apps and mobile technologies. The Evaluation Rubric for Mobile Apps provides a

research-based validated tool that could be a part of a system toolbox to help guide practitioners in the selection and use of Apps in education settings.

Another leadership implication involves the need for professional development for educators. As mobile technologies continue to evolve, leadership teams within school systems and agencies have a responsibility to ensure practitioners are provided with research-based professional development. Best practices for implementing CAI, in particular effective methods to incorporate Apps as part of their instructional program need to be shared. The Evaluation Rubric for Mobile Apps can be a tool to assist in the process of identifying the best Apps to be used in a framework of CAI.

The final implication of this study addresses the possible need to examine the possibility of customization of the Evaluation Rubric for Mobile Apps. Working with SMEs across multiple education fields (e.g., speech and language pathologists, assistive technology specialists, and classroom teachers), unique needs were articulated. For example, some assistive technology specialists cited accessibility for students with disabilities as an important feature of an App. While a number of best practices in teaching and learning are common in all education fields, it is possible revisions might be indicated to capture the unique practices or populations served by certain specialties within the fields of education.

Limitations of the Study

Researchers in the field have questioned the use of the Delphi Method in research. While the Delphi has a number of strengths, Murray & Hammons (1995) noted several factors that need to be considered when employing this method in research. The reliability of the Delphi is dependent on the expertise of the SMEs selected for participation on the panel. The author notes care should be taken to ensure panel members have the experience

and expertise with the matter under consideration in the study.

The rationale for the selection of SMEs for this study is presented in the Participants section. The rationale used to identify SMEs for this investigation satisfies the requirements outlined by the researchers cited above.

While all SMEs met the criteria of expertise, they all volunteered to participate in this investigation. This could be viewed as a limitation as one might predict more favorable responses to the survey data given the participants' interest in the topic and their understanding of the need for such an evaluation tool. The use of a self-report Likert Scale could be a limitation as participants may have related how they would want to be viewed rather than their actual practices.

Another potential disadvantage associated with using the Delphi method is the possibility questions formulated by the researcher may influence panel members. This issue is referred to in the literature as *acquiescence bias* (Knowles & Nathan, 1997). Given the research base supporting the development of the rubric domains, researcher influence did not appear be a factor in this investigation. The first question type in the survey instrument asked participants to rank order research-based practices and not the views or opinions of the researcher in this investigation. The second type of question was open-ended in design and sought additional insights from panel members into the evaluation process of App quality.

The third disadvantage to the Delphi discussed by Murray has to do with the issue of motivation of participants to complete all rounds of the Delphi method. The invitations to be extended to potential participants explicitly asked if they were willing to complete all three rounds in their study. Each round was completed using an online survey, limiting the

amount of time required of participants. All participants who agreed to complete all three rounds were provided access to all study-related data, as well as the final rubric. Interest in the topic and the desire to add to the research base of App use appeared to provide the motivation necessary for a large number of SMEs to participate in the study.

The final limitation of this investigation is the issue of generalizability. While efforts were made to include up to two hundred practitioners from around the world, ninety-four SMEs participated in the round one study, seventy-four in the round two survey, and sixty-five in the follow up survey. Generalization was strengthened by the variety of fields represented and by the fact that all participants were actively engaged using Apps in their various specialties.

Recommendations for Future Research

As a result of this study, practitioners and researchers have an empirically validated instrument to evaluate the quality of mobile Apps aligned with best practices in teaching and learning. A panel of SMEs established high content validity for the Evaluation Rubric for Mobile Apps. Moving forward, experiences using the instrument in the field may indicate the need for revisions. These future iterations may result in even higher content validity. While the process of accumulating evidence to support the validity of test score interpretations starts prior to the development of an assessment (AERA, 1999), Messick (1989) has stated validation is a continuous process. This iterative process will need to continue as the field evolves.

A high degree of SME agreement was obtained through the survey data as to the essential domains of App quality. With high content validity established, a logical next step for research would be to test the reliability of rubric. While high degrees of agreement were obtained as to the language clarity of the score descriptors in each domain, inter-rater

reliability was not evaluated in this study. Moving forward, this would be an important consideration in judging the rubric's utility.

Additional research will need to be conducted to determine how quality Apps might best be incorporated into a program of CAI. As quality Apps are identified, research will be needed to identify what CAI models might be most effective to enhance traditional instruction presented by teachers or practitioners. Questions to guide such research might include, "How much time should a students spend on a App compared to direct instruction?" Another area of investigation might seek to provide research-based guidance to practitioners in terms of how Apps are used. For example, might a given App be best suited for reinforcement, extension, or direct instruction?

Another recommendation for future research would be to determine if different rubrics are more effective for different disciplines in the field of education. For example, are certain domains more critical to evaluating app quality for speech and language pathologists opposed to classroom teachers? Another question worthy of investigation would be, "Are there certain domains that need to be added to measure quality for fields such as assistive technology?"

Data obtained from the surveys and interviews suggested specialization of the rubric might be indicated to fit the needs of particular fields. A search on the Internet using the terms "app rubric" yielded 750,000 references. A number of these references directly cite the instrument developed for this study, the Evaluation Rubric Mobile Apps. Other references cite App rubrics that are revisions of the Evaluation Rubric for Mobile Apps. When the rubric was developed in 2010, there was no documented evidence anyone in the field was examining app quality through the use of a rubric. The plethora of App rubrics

that are coming online is encouraging, but at the same time indicative of the need for additional research.

The Evaluation Rubric for Mobile Apps has had a significant impact on the field of mobile technology in education settings. The researcher's blog, I Teach Therefore iPod, has received over 6,000 page views related to the Evaluation Rubric for Mobile Apps. The rubric and sanctioned revisions can be found on over 50 websites and blogs. The rubric and its iterations are in use in at least 36 states and 12 countries. When this research is published, expanded use of the rubric would be expected. Most importantly, this research has sparked collaboration and inquiry from teachers, practitioners, therapists, and university professors across the globe to examine how Apps are being used in education settings. The validation of the Evaluation Rubric for Mobile Apps provides researchers and practitioners with a framework and common language to begin the process of identifying best practices to leverage these emerging technologies to support teaching and learning.

References

- Adler, M., & Ziglio, E. (1996). *Gazing into the Oracle: The Delphi method and its App to social policy and public health*. London: Kingsley Publishers.
- American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Allan, S. D., & Tomlinson, C. A. (2000). *Leadership for differentiating schools and classrooms*. Alexandria, VA: ASCD.
- Allen, S., & Knight, J. (2009). A method for collaboratively developing and validating a rubric. *International Journal for the Scholarship of Teaching and Learning*, 3(2), 1-17.
- Anastasi, A., & Urbina, A. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Anderson, R. & Dexter, S. (2005). School technology leadership: An empirical investigation of prevalence and effect. *Educational Administration Quarterly*. 41, 49-82.
- App Store Stats Summary. Retrieved September 3, 2011 from <http://148Apps.biz/App-store-metrics/?mpage>
- Arter, J., & McTighe, J. (2001). *Scoring Rubrics in the Classroom*, Corwin: Thousand Oaks, CA.
- Austin, J. D., Hirstein, J., & Walen, S. (1997). Integrated mathematics interfaced with science. *School Science and Mathematics*, 97(1), 45–49.

- Becker, H. J. (1983). How schools use micro-computers. *Classroom Computer Learning* 4(2), 41-44.
- Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, 61, 228-247.
- Bransford, J., Sherwood, R., Hasselbring, T., Kinzer, C., & Williams, S. (1990). Anchored instruction: Why we need it and how technology can help. In D. Nix & R. Spiro (Eds.), *Cognition education and multimedia: Exploring ideas in high technology*. Hillsdale, NJ: Lawrence Erlbaum.
- Brian, M. (2011, July 16). Mobile Apps: A Look at What Makes a Good App Great. [Web blog post]. Retrieved from <http://thenextweb.com/mobile/2011/07/16/mobile-Apps-a-look-at-what-makes-a-good-App-great/>
- Brinkerhoff, J. (2006). Effects of a long-duration, professional development academy on technology, *Journal of Research on Technology in Education*, 39(1), 22-43.
- Brooks, K.W. (1979). Delphi Technique: Expanding Apps. *North Central Association Quarterly*, 53(3), 377-385.
- Brophy, J. E. (1988). Research linking teacher behavior to student achievement: Potential implications for instruction of Chapter I students. *Educational Psychologist*, 23, 235-312.
- Brown, A.L. (1997), "Transforming schools into communities of thinking and learning about serious matters", *American Psychologist*, 5(52), 399-413.
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Newbury Park, CA: Sage Publications.

- Chen, W., Tan, N., Looi, C. K., Zhang, B. H. & Seow, P. (2008). Handheld computers as cognitive tools: technology-enhanced environmental learning. *Research and Practice in Technology-Enabled Learning, World Scientific*, 3(3), 231–252.
- Cheung, A.C.K., & Slavin, R. E. (2103). The effectiveness of educational technology Apps for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88–113.
- Cheung, W. S., & Hew, K.F. (2009). A review of research methodologies used in studies on mobile handheld devices in K-12 and higher education settings. *Australasian Journal of Educational Technology*, 25(2), 153-183.
- Christmann, E. P., & Badgett, J. L. (2000). The comparative effectiveness of CAI on collegiate academic performance. *Journal of Computing in Higher Education*, 11(2), 91-103.
- Cochran, S. W. (1983). The Delphi Method: Formulating and refining group judgments. *Journal of Human Sciences*, 2(2), 111-117.
- Codding R. S., Feinberg, A. B., Dunn, E. K., & Pace, G. M. (2005). Effects of immediate performance feedback on implementation of behavior support plans. *Journal of Applied Behavior Analysis* 38(2), 205-219.
- Cole, H. & Stanton, D. (2003). Designing mobile technologies to support co-present collaboration. *Personal and Ubiquitous Computing*, 7(6), 365-371.
- Cole, N. S. (1990). Conceptions of educational achievement. *Educational Researcher*, 19, 32-37.
- Collins, A. (1988). Cognitive Apprenticeship and instructional technology (Technical Report 6899): BBN Labs Inc., Cambridge, MA.

- Colton, S., & Hatcher, T. (2004). *The web-based Delphi research technique as a method for content validation in HRD and adult education research*. Academy of Human Resource Development International Research Conference, Austin, TX, Feb.29 – Mar 4, 2004. 885-897.
- Cornish, E. (1977). *The study of the future*. World Future Society: Washington, D.C.
- Csikszentmihalyi, M., & Nakamura, J. (1989). The dynamics of intrinsic motivation. In R. Ames & C. Ames (Eds.), *Handbook of motivation theory and research* (Vol. 3). New York: Academic Press.
- Daal, V. & Sandvik, J. M. (2013). The effects of multimedia on literacy development of children at risk: A meta-analysis. *Technology as a Support for Literacy Achievements for Children at Risk*, 7, 73-119.
- Dailey, A. L. (1988). Faculty consensus at a multi-campus college through Delphi. *Community/Junior College Quarterly of Research and Practice*, 12(2), 21-26.
- Delbecq, A. L., Van de Ven, A. H., & Gustafson, D. H. (1975). *Group techniques for program planning*. Glenview, IL: Scott, Foresman, and Company
- Delimarsky, D. (2011, January 12). What makes an App a good App - 7 pointers. [Web blog post]. Retrieved August 28, 2011 from <http://dotnet.dzone.com/articles/what-makes-App-good-App-10>
- Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect. *Journal of Applied Psychology*, 84,795–805.
- Eccles, J. S., & Wigfield, A. (1985). Teacher expectancies and student motivation. In J. B. Dusek (Ed.), *Teacher expectancies*. Hillsdale, NJ: Erlbaum.
- Edwards, J., Norton, S., Taylor, S., Weiss, M. & Van Drusseldrop, R. (1975). How effective is CAI? A review of the research. *Research in Review*, 33, 147-153.

- Ellis, E. S., & Worthington, L. A. (1994). *Research synthesis on effective teaching principles and the design of quality tools for educators* (Technical Report No. 5). Eugene: University of Oregon, National Center to Improve the Tools of Educators.
- Elmer-DeWitt, P. (2011, July 7). How many ways can you say 15 billion App Store downloads? CNN Money. [Web blog post]. Retrieved September 3, 2011, from <http://tech.fortune.cnn.com/2011/07/07/how-many-ways-can-you-say-15-billion-App-store-downloads/>
- Everett, A. (1993). Piercing the veil of the future: A review of the Delphi method of research. *Professional Nurse*, 9(3), 181-185.
- Field, R. (2005). *Favourable conditions for effective and efficient learning in a blended face-to-face/online method*. Proceedings of ASCLITE. Retrieved from http://www.asclite.org.au/conferences/brisbane05/blogs/proceedings/23_Field.pdf
- Flanagan, L. & Jacobsen, M. (2003). Technology leadership for the twenty-first century principal. *Journal of Educational Administration*, 14(2), 124-142.
- Fletcher-Flinn, C. M., & Gravatt, B. (1995). The efficacy of computer assisted instruction (CAI): A meta-analysis. *Journal of Educational Computing Research*, 12(3), 219-242.
- Foxcroft, C.D., Paterson, H., Le Roux, N., & Herbst, D. (2004). *Psychological assessment in South Africa: A needs analysis*. The test use patterns and needs of psychological practitioners (Human Sciences Research Council Rep.). Pretoria: HSRC.
- Frاند, J. (2000). The information-age mindset: Changes in students and implications for higher education. *EDUCAUSE Review*, September-October, 15-24.

- Franklin, T., & Peng, L. (2008) Mobile math: math educators and students engage in mobile learning. *Journal of Computing in Higher Education*, 20(20), 69-80.
- Fuson, K., & Brinko, K. (1985). The comparative effectiveness of microcomputers and flash cards in the drill and practice of basic mathematics facts. *Journal of Research in Mathematics Education*, 5(1), 225-232.
- Gee, J. P. (2008). *Getting over the slump: Innovation strategies to promote children's learning*. New York: The Joan Ganz Cooney Center at Sesame Workshop.
- Global mobile statistics 2012: All quality mobile marketing research, mobile Web stats, subscribers, ad revenue, usage, trends. Retrieved June 12, 2012 from <http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats#subscribers>
- Gros, B. (2003). The impact of digital games in education. *First Monday*, (8)7. Retrieved June 22, 2012 from http://www.firstmonday.org/issues/issue8_7/xyzgros/index.html.
- Halverson, R., & Thomas, C. N. (2007). *The roles and practices of student services staff as data-driven instructional leaders*. In M. Mangin & S. Stoelinga (Eds.), *Instructional teachers leadership roles: Using research to inform and reform* (pp. 163–200). New York: Teachers College Press.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.
- Helmer, O. (1977). Problems in futures research: Delphi and causal cross-impact analysis. *Futures*, February 1977, 17-31.
- Hennessy, S. (2000). Graphing investigations using portable (palmtop) technology. *Journal of Computer Assisted Learning*, 16(3), 243-258.

- Herman, J., & Gribbons, B. (2001). *Lessons learned in using data to support school inquiry and continuous improvement: Final report to the Stuart Foundation*. Los Angeles, CA: University of California, Center for the Study of Evaluation (CSE).
- Herrington, J., Oliver R., & Reeves, T. C. (2003). Patterns of engagement in authentic online learning environments. *Australian Journal of Educational Technology*, 19(1), 59–71.
- Hintzman, D. L. (1974). *Theoretical implications of the spacing effect*. In R. L. Solso (Ed.), *Theories in cognitive psychology*. "The Loyola Symposium (pp. 77-99). Potomac, MD: Erlbaum.
- Howe, N., & Strauss, W. (2001). *Millennials Rising: The Next Great Generation*. New York: Random House.
- Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of Consumer Research*, 30, 138 –149
- Jonassen, D. (1991). Evaluating constructivist learning. *Educational Technology*, 31(9), 28-33.
- Jonsson, A. & Svingby, G. (2007). The use of score rubrics: reliability, validity and educational consequences. *Educational Research Review*, 2, 130-144.
- Kain, D. L. (1993). Cabbages and kings: Research directions in integrated/interdisciplinary curriculum. *The Journal of Educational Thought*, 27(3), 312–331.
- Ke, F. (2008). A qualitative meta-analysis of computer games as learning tools. In R. E. Ferdig (Ed.), *Handbook of Research on Effective Electronic Gaming in Education* (pp. 1-32), New York: IGI Global.

- Knowles, E. S., & Nathan, K. T. (1997). Acquiescent responding in self-reports: Cognitive style or social concern. *Journal of Research in Personality*, 31, 293-301.
- Kulik, J., Bangert-Drowns, R., & Williams, G. (1983). Effects of computer-based instruction on secondary school students. *Journal of Educational Psychology*, 75, 19-26.
- Kulik, J. & Kulik C. (1991). "Effectiveness of computer-based instruction: An updated analysis," *Computers in Human Behavior*, 7(1-2).
- Lachat, M. A., & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Placed At Risk*, 10(3), 333–349.
- Landford, H. W. (1972). *Technological Forecasting Methodologies: A synthesis*. New York: American Management Association, Inc.
- Larson, R. (2000). Toward a psychology of positive youth development. *American Psychologist*, 55(1), 170-183.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563-575.
- Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport*, 59, 277–287.
- Lepper, M. R. (1988). Motivational considerations in the study of instruction. *Cognition and Instruction*, 5, 289-309.
- Leutner, D. (1993). Guided discovery learning with computer-based simulation games: Effects of adaptive and non-adaptive instructional support. 3(2), 113-132.

- Levien, R. E. (1972). *The emerging technology: Instructional uses of the computer in higher education*, New York, NY: McGraw-Hill Book Company.
- Lindeman, C. (1975). Delphi survey of priorities in clinical nursing research. *Nursing Research*, 24(6), 434-441.
- Looi, C. K., Seow, P., Zhang, B., So, H. J., Chen, W-L., & Wong, L. H. (2010). Leveraging mobile technology for sustainable seamless learning. *British Journal of Educational Technology*. 41(2), 154-169.
- Lowe, J. (2001). Computer-based education: Is it a panacea? *Journal of Research on Technology in Education*, 34(2), 163-171.
- Maehr, M. L. (1984). *Meaning and motivation: Toward a theory of personal investment*. In R. Ames, & C. Ames (Eds.), *Research on motivation in education* (pp. 115-144). New York: Academic Press.
- Malini, Y. R. (2011). Design and development of rubrics to improve assessment outcomes: A pilot study in a Master's level business program in India. *Quality Assurance in Education*, 19(1), 88-104.
- Malone, T. (1980). What makes things fun to learn? A study of intrinsically motivating computer games. *Cognitive and Instructional Sciences Series* (Research Report No. CIS-7 SSL-80-11). Palo Alto, CA: Palo Alto Research Center.
- Mandinach, E.B. (1987). Clarifying the 'A' in CAI for learners of different abilities. *Journal of Educational Computing Research*, 3(1), 113-128.
- Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.

- McNamara, F. (1996). *Measuring second language performance*. London: Longman.
- Meech, S. (2011, March 18). Purpose of IEAR.org. [Web blog post]. Retrieved August 28, 2011 from <http://www.ear.org/ear/2011/3/18/purpose-of-earorg.html>
- Messick, S. (1989). Validity. In R. L. Linn (ed.), *Educational Measurement* (3rd ed.). New York: Macmillan.
- Molnar, A. R. (1975). Viable goals for new educational technology efforts: Science education and the new technological revolution, *Educational Technology*, 15(9).
- Moss, V. D. (1996). *The efficacy of massed versus distributed practice as a function of desired learning outcomes and grade level of the student*. (Doctoral dissertation, Utah State University, 1995). Dissertation Abstracts International, 56, 5204.
- TNW Mobile. (2011, July 17). *The Next Web - International technology news, business & culture*. Mobile Apps: A look at what makes an App popular. [Web blog post]. Retrieved September 3, 2011, from <http://thenextweb.com/mobile/2011/07/16/mobile-Apps-a-look-at-what-makes-a-good-App-great/>
- Murphy, M. K., Black, N., Lamping, D. L. McKee, C. M., Sanderson, C. F. B., Askham, J. (1988). Consensus development methods and their use in clinical guideline development. *Health Technology Assessment*, 2(3).
- Murray, J. W., & Hammons, J.O. (1995). Delphi: A versatile methodology for conducting qualitative research. *The Review of Higher Education*, 18 (4), 423-436.
- Nagel, D. (2012), App sells twice as many iPads as Macs to school. Retrieved January 4, 2013 from <http://thejournal.com/articles/2012/07/25/Apple-sells-twice-as-many-ipads-as-macs-to-u.s.-schools.aspx>

- Newmann, F., Secada, W., & Wehlage, G. (1995). *A guide to authentic instruction and assessment: Vision, standards and scoring*. Alexandria, VA: ASCD.
- Nielsen, J., & Hackos, J. T. (1993). *Usability engineering, (125184069)*. Boston: Academic Press.
- Norman, D. (1993). *Things that make us smart: Defending human attributes in the age of the machine*. Reading, MA: Addison-Wesley.
- NPD Group. (2008). *Kids and consumer electronics IV*. Port Washington, NY: NPD Group.
- Nunnally, J. C., (1967). *Psychometric theory*. New York: McGraw-Hill.
- Oblinger, D. (2003). Boomers, Gen-Xers, and Millennials: Understanding the new students. *EDUCAUSE Review*, 38(4), 37-47.
- Office of Technology Assessment, (1982). *Informational technology and its impact on American education*. U.S. Congress, Washington, DC, pp.128-133.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Pintrich, P. R., & Schunk, D. (2002). *Motivation in education: Theory, research, and Apps* (2nd ed.). Upper Saddle, NJ: Prentice-Hall, Inc.
- Pintrich, P.R., Smith D.A.F., Garcia T. & McKeachie W.J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ), *Educational and Psychological Measurement*, 53, 801-803.
- Popham, W.J. (1997). What's wrong and what's right with rubrics? *Educational Leadership*, 55(2), 72-5.
- Prensky, M. (2001). Digital Natives, Digital Immigrants Part 1, *On the Horizon*, 9(5), 1-6.

- Pugh, K. J., & Bergin, D. A. (2006). Motivational influences on transfer. *Educational Psychologist, 41*(3), 147–160.
- Quinn, R. E. (1996). *Deep change: Discovering the leader within*. San Francisco: Jossey-Bass
- Resnick, L. B. (1987). Learning in and out of school. 1987 AERA presidential address. *Educational Research, 13*-20.
- Risden, K., Hanna, E., & Kanerva, A. (1997, April). *Dimensions of intrinsic motivation in children's favorite computer activities*. Poster session presented at the meeting of the Society for Research in Child Development, Washington, DC
- Roschelle, J. (2007). *Scientifically-Based Research Studies Examining the Use of Technology in Mathematics Education*. Presentation at the Consortium for School Networking (CoSN), 12th Annual K-12 School Networking Conference, San Francisco, CA.
- Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting, 15*(4), 353-375.
- Rowinski, D. (2012). History of mobile app stores. Retrieved May 4, 2013, from https://2012/02/06/infographic_history_of_mobile_App_stores
- Ryan, A. W. (1991). Meta-analysis of achievement effects in microcomputer applications in elementary schools. *Educational Administration Quarterly, 27*(2), 161-184.
- Sackman H. (1975). *Delphi Critique*. Lexington Books, Boston, MA

- Salkind, N. J. (2007). *Statistics for People Who (Think They) Hate Statistics: Excel 2010 Edition*. Thousand Oaks, CA: Sage
- Sandford, R., Ulicsak, M., Facer, K., & Rudd, T. (2007). Teaching with games. *Learning, Media & Technology*, 32 (1), 101-105.
- Scheeler, M. C., Ruhl, K. L., & McAfee, J. K. (2004). Providing performance feedback to teachers: A review. *Teacher Education and Special Education*, 27(4), 396-407.
- Shuler, C. (2012). iLearn II: An analysis of the education category of Apple's App store. New York: The Joan Ganz Cooney Center at Sesame Workshop. Retrieved April 4, 2013 from <http://www.joanganzcooneycenter.org/wp-content/uploads/2012/01/ilearnii.pdf>
- Shuler, C. (2009). *Pockets of Potential" Using mobile technologies to promote children's learning*. New York: The Joan Ganz Cooney Center at Sesame Workshop. Retrieved July 18, 2012 from http://www.joanganzcooneycenter.org/pdf/pockets_of_potential.pdf
- Skulmoski, G., & Hartman, F. (2007). The Delphi method: Researching what does not exist (yet). *Proceedings of the International Research Network on Organization*
- Soloway, E. Norris, C., Blumenfeld, P., Fishman, B. J., & Marx, R. (2001). Devices are ready-at-hand. *Communications of the ACM*, 44(6), 15-20.
- Spikol, D., Milrad, M., Maldonado, H. and Pea, R. (2009). Integrating co-design practices into the development of mobile science collaboratories. Proceedings of the 9th IEEE International Conference on Advanced Learning Technologies (ICALT 2009), July 14-18, 2009. Riga, Latvia.

Squire, K., & Klopfer, E. (2007). Augmented reality simulations on handheld computers.

Journal of the Learning Sciences, 16(3), 371 - 413.

Steel Media Network (2013). Apple store metrics. Retrieved on May 4, 2013 from

<http://148Apps.bix?App-store-metrics/?mpage>

Strauss, W., & Howe, N. (1991). *Generations: The History of America's Future, 1584-*

2069. New York: Quill.

Suppes, P. (1980). The computer in the school: Tutor, tool, tutee. In Robert T. Taylor, (Ed.)

pp. 213-260, New York, NY: Teachers College Press.

Tomlinson, C. (1999). *The Differentiated Classroom: Responding to the Needs of All*

Learners. Alexandria, VA: Association for Supervision and Curriculum

Development.

Tomlinson, C. (2003). *Fulfilling the Promise of the Differentiated Classroom: Strategies*

and Tools for Responsive Teaching. Alexandria, VA: Association for Supervision

and Curriculum Development.

Underwood, B. J. (1970). A breakdown of the total-time law in free-recall learning.

Journal of Verbal Learning and Verbal Behavior, 9, 573-580.

U.S. Department of Education, National Center for Education Statistics. (2010). Teachers'

use of educational technology in U.S. public schools: 2009. (NCES 2010-040).

Vahey, P., & Crawford, V. (2002). *Palm Education Pioneers Program: Final evaluation*

report. Menlo Park, CA: SRI International.

- Van Daal, V., & Sandvik, J. (2011). Reading research and the promotion of reading outside of school: A comprehensive review of empirical research. *International Journal About Parents in Education*, 5(2), 110-122.
- Van't Hoofvt, M., Diaz, S. & Swan, K. (2004). Examining the potential of handheld computers: Findings from the Ohio PEP project. *Journal of Educational Computing Research*, 30(4), 295-312.
- Vavoula, G., & Sharples, M. (2009). Meeting the Challenges in Evaluating Mobile Learning: A 3-level Evaluation Framework. *International Journal of Mobile and Blended Learning*, 1(2), pp. 54-75.
- Vogel, J.F., Vogel, D.S., Cannon-Bowers, J., Bowers, C.A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, 34(3), 229-243.
- Vygotsky, L. S., (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Young, M.F. (1993). Instructional design for situated learning. *Educational Technology Research and Development*, 41(1), 43-58.
- Westbrook, L. (1997). Information access issues for interdisciplinary scholars: Results of a Delphi study on women's studies research. *The Journal of Academic Librarianship*, 23 (May), 211-216.
- Whalen, S., & Csikszentmihalyi, M. (1991). *Putting flow into educational practice*. Chicago, IL: University of Chicago.
- Whitehead, A. (1932). *The Aims of Education and Other Essays*, Williams & Norgate, London.

- Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review*, 6, 49–78.
- Winn, W. (1993). Instructional design and situated learning: Paradox or partnership. *Educational Technology*, 33(3), 16-21.
- Wouters, P., van Nimwegen, C., van Oostendorp, H. & van der Spek, E.D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 105 (1).

Appendix A
Sample of Delphi Research (Skulmoski et al (2007))

Study	Delphi Focus	Rounds	Sample Size
Gustafson, Shukla, Delbecq, & Walster (1973)	Estimate almanac events to investigate Delphi accuracy	2	4
Hartman & Baldwin (1995)	Validate research outcomes	1	62
Czinkota & Ronkainen (1997)	Impact analysis of changes to the International business environment.	3	34
Kuo & Yu (1999)	Identify national park selection criteria.	1	28
Nambisan et al. (1999)	Develop taxonomy of organizational mechanisms.	3	6
Lam, Petri, & Smith (2000)	Develop rules for a ceramic casting process.	3	3
Roberson, Collins, & Oreg (2005)	Examine and explain how recruitment message specificity influences job seeker attraction to organizations.	2	171
Niederman, Brancheau, & Wetherbe, (1991)	Survey senior IS executives to determine the most critical IS issues for the 1990s.	3	114, 126 & 104
Duncan (1995)	Identify and rank the critical elements of IS infrastructure flexibility.	2	21
Brancheau, Janz, & Wetherbe (1996)	Survey SIM members to determine the most critical IS issues for the near future,	3	78, 87 & 76
Nambisan et al. (1999)	Develop a taxonomy of knowledge creation mechanisms.	3	11
Scott (2000)	Rank technology management issues in new product development	3	20
Wynekoop & Walz (2000)	Rank the most important characteristics of high performing IT personnel.	3	9
R. Schmidt, Lyytinen, Keil, & Cule (2001)	Identify and rank software development project risks: an international comparative study.	3	Finland – 13 Hong Kong- 11 U.S. - 21
Keil, Tiwana, & Bush (2002)	Rank software development project risks.	3	15, 15 & 10

Appendix B Demographic Data of SMEs

Key for demographic information

SME Subgroup	Abbreviation
University affiliated	U
Students	S
Speech & Language	SLP
Administrators	Admin
Assistive Technology	AT
Library Media	LMS
Technology Directors/Support	Tech
Teachers	T

Demographic	One Year	Two Years	Three Years	More than three
Experience with Mobile Technology	U – 2 S – 1 SLP – 1 Admin – 3 AT – 1 LMS – 2 Tech – 4 T – 4 N - 18 (19.1%)	U – 3 S – 2 SLP – 0 Admin – 2 AT – 2 LMS – 0 Tech – 3 T – 4 N - 16 (17.0%)	U – 5 S – 0 SLP – 7 Admin – 5 AT – 1 LMS – 0 Tech – 3 T – 5 N - 26 (27.7%)	U – 6 S – 3 SLP – 3 Admin – 4 AT – 7 LMS – 0 Tech – 7 T – 4 N - 34 (36.2%)
	Not as advanced	About the same	More advanced	Greatly advanced
Experience relative to other in the field	U – 1 S – 0 SLP – 0 Admin – 1 AT – 0 LMS – 0 Tech – 0 T – 0 N - 2 (2.2%)	U – 3 S – 1 SLP – 1 Admin – 2 AT – 1 LMS – 0 Tech – 2 T – 2 N - 12 (12.8%)	U – 6 S – 1 SLP – 7 Admin – 9 AT – 6 LMS – 2 Tech – 11 T – 10 N - 52 (55.3%)	U – 6 S – 4 SLP – 3 Admin – 2 AT – 4 LMS – 0 Tech – 4 T – 5 N - 28 (29.8%)
	Bachelors	Masters	Doctorate	
Level of Education	U – 0 S – 1 SLP – 0 Admin – 4 AT – 3 LMS – 0 Tech – 1 T – 4 N - 13 (13.8%)	U – 5 S – 4 SLP – 11 Admin – 9 AT – 8 LMS – 2 Tech – 15 T – 12 N - 66 (70.2%)	U – 11 S – 1 SLP – 0 Admin – 1 AT – 0 LMS – 0 Tech – 1 T – 1 N - 15 (16.0%)	

	Yes	No		
Influence in decision making regarding App purchases	U – 12 S – 6 SLP – 10 Admin – 14 AT – 11 LMS – 2 Tech – 16 T – 14 N - 85 (90.4%)	U – 4 S – 0 SLP – 1 Admin – 0 AT – 0 LMS – 0 Tech – 1 T – 3 N - 13 (9.6%)		

Appendix C
Original Evaluation Rubric for Mobile Apps (developed 10/18/2010)

Evaluation Rubric for iPod Apps

Domain	1	2	3	4
Curriculum Connection	Skill(s) reinforced in the app are not clearly connected to the targeted skill or concept	Skill(s) reinforced are prerequisite or foundation skills for the targeted skill or concept	Skill(s) reinforced are related to the targeted skill or concept	Skill(s) reinforced are strongly connected to the targeted skill or concept
Authenticity	Skills are practiced in a rote or isolated fashion (e.g., flashcards)	Skills are practiced in a contrived game/simulation format	Some aspects of the app are presented in an authentic learning environment	Targeted skills are practiced in an authentic format/problem-based learning environment
Feedback	Feedback is limited to correctness of student responses	Feedback is limited to correctness of student responses and may allow for student to try again	Feedback is specific and results in improved student performance (may include tutorial aids)	Feedback is specific and results in improved student performance; Data is available electronically to student and teacher
Differentiation	App offers no flexibility (settings cannot be altered)	App offers limited flexibility (e.g., few levels such as easy, medium, hard)	App offers more than one degree of flexibility to adjust settings to meet student needs	App offers complete flexibility to alter settings to meet student needs
User Friendliness	Students need constant teacher supervision in order to use the app	Students need to have the teacher review how to use the app on more than one occasion	Students need to have the teacher review how to use the app	Students can launch and navigate within the app independently
Student Motivation	Students avoid the use of the app or complain when the app is assigned by the teacher	Students view the app as "more schoolwork" and may be off-task when directed by the teacher to use the app	Students will use the app as directed by the teacher	Students are highly motivated to use the app and select it as their first choice from a selection of related choices of apps

Created by Harry Walker – Johns Hopkins University 10/18/2010
Please contact for permission to use hwalker@bcps.org

Evaluation Rubric for Apps Survey

This dissertation investigation is being conducted to establish both content and construct validity for the Evaluation Rubric for Mobile Apps. Please read the following consent form carefully.

Your input in the validation process of this rubric is important and will assist in empirically validating this tool, thereby greatly contributing to the educational technology field. Survey responses will be transmitted directly to the principal investigator from an online survey and all responses will remain confidential and totally anonymous. You will not be asked to provide any personal information in your responses. Your responses cannot be traced back to you. Your participation is entirely voluntary, and you do not have to participate.

If you choose to participate, please download a copy of the Evaluation Rubric for Mobile Apps that accompanied your invitation email. This will ensure you have the most recently revised rubric to use when completing the survey.

If you wish to participate in this process to validate the rubric, your consent will be made by clicking on the first button, "I Consent" below and then the "Next" button at the bottom of the page; this will continue the survey. If you do not wish to participate, click on the second button, "I DO NOT consent" then the next button and the survey will be terminated. Please respond below:

1. Do you wish to participate in this investigation?

☐ I Consent: I understand the information presented above. I consent to participate in this investigation.

☐ I DO NOT Consent: I understand the information presented above. I do not consent to participate in this investigation.

Purpose of Research

As you aware, evaluating the quality of a mobile App (App) is a subjective process. The goal of this Delphi research project is to empirically validate a rubric practitioners can use to objectively evaluate mobile Apps. Your feedback on the rubric is critical in this process. While the rubric has been widely adopted by school systems, universities, and agencies around the world, this study will ensure its content and construct validity.

As a participant you will be asked to complete three surveys over the course of two months. The first survey asks for your feedback on how essential each of the domains/areas contained in the rubric is in the process of determining App quality. You are also asked to provide feedback on the wording of the score descriptors within each domain/area in the rubric. In subsequent surveys you will be provided with the data generated by practitioners such as yourself from across the globe. Based on the responses from the group, you will be asked to reflect on your feedback and determine if you would like to revise your ratings/feedback.

Throughout the process, as well as at the conclusion of study, you will be provided with all data generated from the group, as well as the final validated rubric.

The rubric has been revised since it was first posted in October of 2010. The revised rubric you will be evaluating is sent as an attachment to your email invitation. It is marked "Revised 6/25/2012." Please use this revised rubric in your survey responses.

Thank you again for agreeing to participate in this important project. You will receive additional email/surveys in the coming weeks.

2. How well do you think the following domains/areas are necessary and sufficient to address the quality of a mobile App (APP)?

	Poor measure of App quality	Not a good measure of App quality	OK measure of App quality	Good measure of App quality	Excellent measure of App quality
Curriculum Connection	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Authenticity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Feedback	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Differentiation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
User Friendliness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student Use	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Poor measure of App quality	Not a good measure of App quality	OK measure of App quality	Good measure of App quality	Excellent measure of App quality
Student Performance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Should any domains/areas be added or deleted from the rubric? Please explain.

3. How well do you think the point value descriptions differentiate the quality of the levels within each of the domains?

	Descriptions do not differentiate the quality of the levels within each domain/area	Descriptions minimally differentiate the quality of the levels within each domain/area	Descriptions adequately differentiate the quality of the levels within each domain/area	Descriptions strongly differentiate the quality of the levels within each domain/area
Curriculum Connection	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Authenticity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Feedback	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Differentiation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
User Friendliness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student Use	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student Performance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Could the wording of any of the score descriptors be improved to better differentiate the quality of the levels within any domain? If so, which domain and what improvements would you suggest?

4. How many years of experience do you have using mobile technology Apps in your setting?

- ☐ One year
- ☐ Two years
- ☐ Three years
- ☐ More than three years

5. Compared to others in your field/profession, how does your experience/expertise with using Mobile Computing Apps (Apps) compare?

- ☐ Greatly advanced
- ☐ More advanced
- ☐ About the same
- ☐ Not as advanced

6. What is your highest level of education?

- ☐ Bachelor's degree
- ☐ Master's degree
- ☐ Doctoral degree

7. Do you serve in a position that influences decision-making regarding App purchases for your school/agency?

- ☐ Yes
- ☐ No

Thank you for completing this survey. Your input will make a valuable contribution to the field.

App Rubric Survey Round Two

To everyone who completed the first round survey, thank you very much. Your time and expertise are contributing to our understanding of how to most effectively use technology in educational settings. If you WERE NOT able to complete the first round, please click on the "Opt Out" button below. If you wish to continue participating in this Delphi method to further increase the content and construct validity for the "Evaluation Rubric for Mobile Apps", please click on the "Continue" button below. Thank you in advance for your participation.

1. I did not complete the first round survey and wish to "opt out" of this research study.

- ☐ "Opt out"

2. I wish to continue my participation in this research study.

- ☐ Continue

3. Please use the email attachment entitled, "Round One Survey Results" (Table 2) to answer survey questions 3-9.

Based on the summary of ratings by the participants who completed the first round survey, do you wish to change your rating for the domain, Curriculum Connections?

- ☐ No change in my rating
- ☐ Increase my rating
- ☐ Decrease my rating

If you wish to change your rating, please fill in the numbers to the following prompt: I wish to change my rating from _____ to _____

4. Based on the summary of ratings by the participants who completed the first round survey, do you wish to change your rating for the domain, Authenticity?

- ☐ No change in my rating
- ☐ Increase my rating
- ☐ Decrease my rating

If you wish to change your rating, please fill in the numbers to the following prompt: I wish to change my rating from _____ to _____

5. Based on the summary of ratings by the participants who completed the first round survey, do you wish to change your rating for the domain, Feedback?

- ☐ No change in my rating
- ☐ Increase my rating
- ☐ Decrease my rating

If you wish to change your rating, please fill in the numbers to the following prompt: I wish to change my rating from _____ to _____

6. Based on the summary of ratings by the participants who completed the first round survey, do you wish to change your rating for the domain, Differentiation?

- ☐ No change in my rating
- ☐ Increase my rating
- ☐ Decrease my rating

If you wish to change your rating, please fill in the numbers to the following prompt: I wish to change my rating from _____ to _____

7. Based on the summary of ratings by the participants who completed the first round survey, do you wish to change your rating for the domain, User Friendliness?

- ☐ No change in my rating
- ☐ Increase my rating
- ☐ Decrease my rating

If you wish to change your rating, please fill in the numbers to the following prompt: I wish to change my rating from _____ to _____

8. Based on the summary of ratings by the participants who completed the first round survey, do you wish to change your rating for the domain, Student Use?

- ☐ No change in my rating
- ☐ Increase my rating
- ☐ Decrease my rating

If you wish to change your rating, please fill in the numbers to the following prompt: I wish to change my rating from _____ to _____

9. Based on the summary of ratings by the participants who completed the first round survey, do you wish to change your rating for the domain, Student Performance?

- ☐ No change in my rating
- ☐ Increase my rating
- ☐ Decrease my rating

If you wish to change your rating, please fill in the numbers to the following prompt: I wish to change my rating from _____ to _____

10. Does the term "Motivation" better describe the domain currently labeled "Student Use"?

- ☐ Yes

- No

11. Please use the email attachment entitled "Score Description Revisions" to answer the last two questions in this survey.

Participants indicated difficulty differentiating between scores of 2 and 3 in the Curriculum Connection domain. Which matrix of score descriptions in the attachment provides greater clarity?

- Original version of score descriptions
- Revised version of score descriptions

12. Participants indicated difficulty differentiating between scores of 3 and 4 in the Feedback domain. Which matrix of score descriptions in the attachment provides greater clarity?

- Original version of score descriptions
- Revised version of score descriptions

Thank you for completing this second survey in the Delphi method Research Study to increase the content and construct validity for the "Evaluation Rubric for Mobile Apps." Survey results will once again be provided to all participants after the data has been compiled and analyzed.

App Rubric Follow Up Survey

Thank you for agreeing to complete this short online survey as a follow-up to my dissertation research on the Evaluation Rubric for Mobile Apps. The survey should take five minutes or less to complete. The purpose of this survey is to gauge the use of the rubric in current practice and to determine which of the domains have proven to be the most beneficial in making decisions about App purchases. Thank you for continuing to support this important research.

1. Have you used the Evaluation Rubric for Mobile Apps?

- ☐ Yes
- ☐ No

2. If yes, how many times have you used the Evaluation Rubric for Mobile Apps?

- ☐ One time
- ☐ 2-5Times
- ☐ 5-10Times
- ☐ More than 10 times

3. Would you recommend this rubric to colleagues who are interested in evaluating Apps?

- ☐ Highly recommend
- ☐ Recommend
- ☐ Recommend with reservations
- ☐ Not recommend

4. Rank order each of the domains in the rubric in terms of how important they are in determining the quality of an App. (1 the most important, 7 the least important)

			Curriculum Connections
		▼	Authenticity
		▼	Feedback
		▼	Differentiation

		▼	User Friendliness
		▼	Motivation
		▼	Student Performance

5. How are you currently using Apps in your setting?

6. What problems have you encountered in using Apps in your setting?

7. What do you see in the future for App use in your setting?

Thank you for your contributions, supporting the research base to identify best practices in the use of mobile technologies in educational settings. I will send the results of this survey when the data is compiled and analyzed.

1. Have you used the Evaluation Rubric for Mobile iPod Apps to help in your decision making about which Apps to purchase?
2. If so, how many times have you used the rubric?
3. Would you recommend the rubric to a colleague?
4. Which domains are the most important in making decisions about Apps?
5. What other processes/procedures are in place for App purchases in your setting?
6. What types of Apps have you evaluated?
7. How are you currently using Apps in your setting?
8. What are the administrative issues related to using Apps in your setting?
9. What are the instructional implications related to using Apps in your setting?
10. What problems have you encountered in using Apps in your setting?
11. What do you see in the future for App use in your setting?

*Interviews were conducted by telephone to a representative sample of SMEs

Appendix H

Rubric Adoption Prior to Delphi Study

State/Country	Districts/Agencies		State/Country	Districts/Agencies
Alaska	2		Montana	18
Arizona	3		Nebraska	2
Arkansas	1		Nevada	1
California	8		New Jersey	3
Colorado	4		New York	6
Delaware	3		North Dakota	1
Florida	3		Ohio	7
Georgia	5		Oklahoma	2
Illinois	14		Pennsylvania	8
Indiana	2		South Carolina	2
Iowa	6		Tennessee	2
Kansas	2		Texas	8
Maine	1		Utah	2
Maryland	3		Virginia	6
Massachusetts	5		Washington	1
Michigan	6		Washington, DC	2
Minnesota	23		Wisconsin	14
Missouri	2			
Canada	11		Sweden	2

England	6		Taiwan	1
New Zealand	1		Thailand	1
Qatar	1		Australia	8

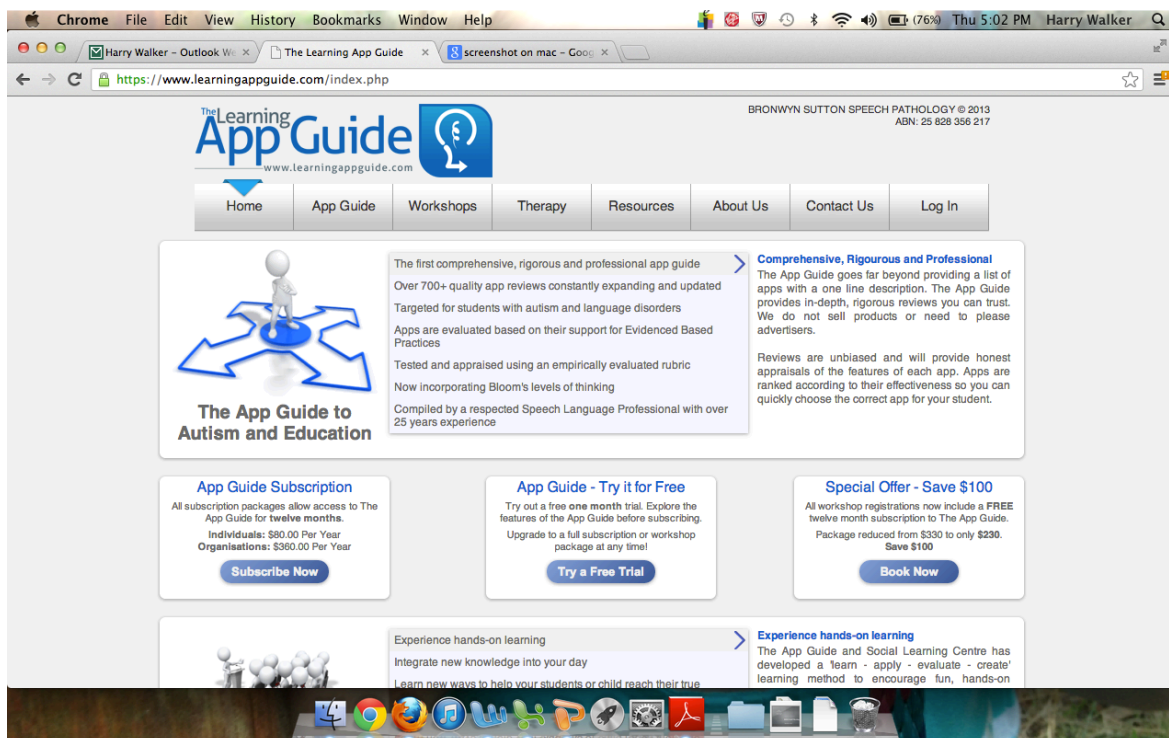
Appendix I
Validated Evaluation Rubric for Mobile Apps

Domain	1	2	3	4
Curriculum Connection	Skill(s) reinforced in the app are not clearly connected to the targeted skill or concept	Skill(s) reinforced are prerequisite or foundation skills for the targeted skill or concept	Skill(s) reinforced are related to the targeted skill or concept	Skill(s) reinforced are strongly connected to the targeted skill or concept
Authenticity	Skills are practiced in a rote or isolated fashion (e.g., flashcards)	Skills are practiced in a contrived game/simulation format	Some aspects of the app are presented in an authentic learning environment	Targeted skills are practiced in an authentic format/problem-based learning environment
Feedback	Feedback is limited to correctness of student responses	Feedback is limited to correctness of student responses & may allow for student to try again	Feedback is specific and results in improved student performance (may include tutorial aids)	Feedback is specific & results in improved student performance; Data is available electronically to student & teacher
Differentiation	App offers no flexibility (settings cannot be altered)	App offers limited flexibility (e.g., few levels such as easy, medium, hard)	App offers more than one degree of flexibility to adjust settings to meet student needs	App offers complete flexibility to alter settings to meet student needs
User Friendliness	Students need constant teacher supervision in order to use the app	Students need to have the teacher review how to use the app on more than one occasion	Students need to have the teacher review how to use the app	Students can launch and navigate within the app independently
Student Use	Students avoid the use of the app or complain when the app is assigned by the teacher	Students view the app as “more schoolwork” and may be off-task when directed by the teacher to use the app	Students will use the app as directed by the teacher	Students are highly motivated to use the app and select it as their first choice from a selection of related apps

Student Performance	Students show no evidence of improved performance as a result of using the app	Students show minimal improvements in performance as a result of using the app	Students show satisfactory improvements in performance as a result of using the app	Students show outstanding improvements in performance as a result of using the app
----------------------------	--	--	---	--

Appendix J

Website Currently Using the App Rubric



Dr. Walker's Biographical Sketch

Dr. Harry Walker was granted his bachelor's degree in Early Childhood, Elementary, and Special Education from Tusculum College in Greeneville Tennessee. The Johns Hopkins University in Baltimore Maryland conferred his Master of Science Degree and Doctorate of Education. His doctoral major was Leadership and Teacher Development.

Dr. Walker worked for 32 years in the Baltimore County Public School System, 15 of which was in the role of Elementary Level Principal in two, Title I schools. After retiring from that school system, he served as Senior Education Consultant for the technology consulting firm, Education Elements, advising schools and school systems across the country on the development and implementation of blended learning classrooms. He is now back in public education, currently serving as Principal of Bellows Spring Elementary School in Howard County, Maryland.

Dr. Walker's interests in education include the effective integration of technology in classrooms and metacognition strategy instruction.